WO0164922

Publication Title:

HETEROLOGOUS EXPRESSION OF NEISSERIAL PROTEINS

Abstract:

Abstract of WO0164922

Alternative and improved approaches to the heterologous expression of the proteins of <i>Neisseria meningitidis</i> and <i>Neisseria gonorrhoeae</i>. These approaches typically affect the level of expression, the ease of purification, the cellular localisation, and/or the immunological properties of the expressed protein. Data supplied from the esp@cenet database - Worldwide

---

120+1

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
7 September 2001 (07.09.2001)

PCT

(10) International Publication Number
WO 01/64922 A2

(51) International Patent Classification[7]: C12N 15/70, C07K 14/22, 19/00

(21) International Application Number: PCT/IB01/00452

(22) International Filing Date: 28 February 2001 (28.02.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
0004695.3 28 February 2000 (28.02.2000) GB
0027675.8 13 November 2000 (13.11.2000) GB

(71) Applicant *(for all designated States except US)*: **CHIRON SPA** [IT/IT]; Via Fiorentina 1, I-53100 Siena (IT).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **ARICO, Maria, Beatrice** [IT/IT]; Chiron SpA, Via Fiorentina 1, I-53100 Siena (IT). **COMANDUCCI, Maurizio** [IT/IT]; Chiron SpA, Via Fiorentina 1, I-53100 Siena (IT). **GALEOTTI, Cesira** [IT/IT]; Chiron SpA, Via Fiorentina 1, I-53100 Siena (IT). **MASIGNANI, Vega** [IT/IT]; Chiron SpA, Via Fiorentina 1, I-53100 Siena (IT). **GUILIANI, Marzia, Monica** [IT/IT]; Chiron SpA, Via Fiorentina 1, I-53100 Siena (IT). **PIZZA, Mariagrazia** [IT/IT]; Chiron SpA, Via Fiorentina, 1, I-53100 Siena (IT).

(74) Agents: **HALLYBONE, Huw, George** et al.; Carpmaels & Ransford, 43 Bloomsbury Square, London WC1A 2RA (GB).

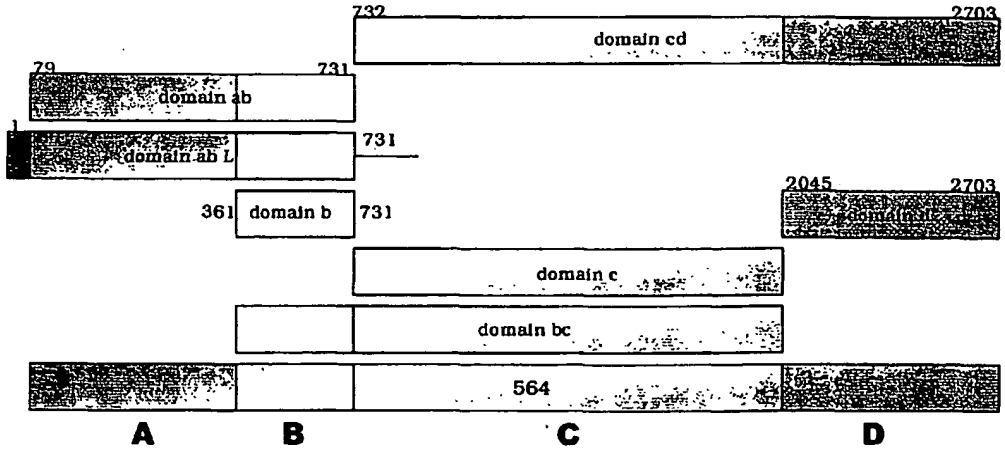
(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: HETEROLOGOUS EXPRESSION OF NEISSERIAL PROTEINS



(57) Abstract: Alternative and improved approaches to the heterologous expression of the proteins of *Neisseria meningitidis* and *Neisseria gonorrhoeae*. These approaches typically affect the level of expression, the ease of purification, the cellular localisation, and/or the immunological properties of the expressed protein.

WO 01/64922 A2

HETEROLOGOUS EXPRESSION OF NEISSERIAL PROTEINS

All documents cited herein are incorporated by reference in their entirety.

TECHNICAL FIELD

This invention is in the field of protein expression. In particular, it relates to the heterologous expression of proteins from *Neisseria* (*e.g. N.gonorrhoeae* or, preferably, *N.meningitidis*).

BACKGROUND ART

International patent applications WO99/24578, WO99/36544, WO99/57280 and WO00/22430 disclose proteins from *Neisseria meningitidis* and *Neisseria gonorrhoeae*. These proteins are typically described as being expressed in *E.coli* (*i.e.* heterologous expression) as either N-terminal GST-fusions or C-terminal His-tag fusions, although other expression systems, including expression in native *Neisseria*, are also disclosed.

It is an object of the present invention to provide alternative and improved approaches for the heterologous expression of these proteins. These approaches will typically affect the level of expression, the ease of purification, the cellular localisation of expression, and/or the immunological properties of the expressed protein.

DISCLOSURE OF THE INVENTION

Nomenclature herein

The 2166 protein sequences disclosed in WO99/24578, WO99/36544 and WO99/57280 are referred to herein by the following SEQ# numbers:

Application	Protein sequences	SEQ# herein
WO99/24578	Even SEQ IDs 2-892	SEQ#s 1-446
WO99/36544	Even SEQ IDs 2-90	SEQ#s 447-491
WO99/57280	Even SEQ IDs 2-3020	SEQ#s 492-2001
	Even SEQ IDs 3040-3114	SEQ#s 2002-2039
	SEQ IDs 3115-3241	SEQ#s 2040-2166

In addition to this SEQ# numbering, the naming conventions used in WO99/24578, WO99/36544 and WO99/57280 are also used (*e.g.* 'ORF4', 'ORF40', 'ORF40-1' *etc.* as used in WO99/24578 and WO99/36544; 'm919', 'g919' and 'a919' *etc.* as used in WO99/57280).

The 2160 proteins NMB0001 to NMB2160 from Tettelin *et al.* [*Science* (2000) 287:1809-1815] are referred to herein as SEQ#s 2167-4326 [see also WO00/66791].

The term 'protein of the invention' as used herein refers to a protein comprising:

(a) one of sequences SEQ#s 1-4326; or

(b) a sequence having sequence identity to one of SEQ#s 1-4326; or

(c) a fragment of one of SEQ#s 1-4326.

The degree of 'sequence identity' referred to in (b) is preferably greater than 50% (*eg.* 60%, 70%, 80%, 90%, 95%, 99% or more). This includes mutants and allelic variants [*e.g.* see WO00/66741]. Identity is preferably determined by the Smith-Waterman homology search algorithm as implemented in the MPSRCH program (Oxford Molecular), using an affine gap search with parameters *gap open penalty=12* and *gap extension penalty=1*. Typically, 50% identity or more between two proteins is considered to be an indication of functional equivalence.

The 'fragment' referred to in (c) should comprise at least *n* consecutive amino acids from one of SEQ#s 1-4326 and, depending on the particular sequence, *n* is 7 or more (*eg.* 8, 10, 12, 14, 16, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100 or more). Preferably the fragment comprises an epitope from one of SEQ#s 1-4326. Preferred fragments are those disclosed in WO00/71574 and WO01/04316.

Preferred proteins of the invention are found in *N.meningitidis* serogroup B.

Preferred proteins for use according to the invention are those of serogroup B *N.meningitidis* strain 2996 or strain 394/98 (a New Zealand strain). Unless otherwise stated, proteins mentioned herein are from *N.meningitidis* strain 2996. It will be appreciated, however, that the invention is not in general limited by strain. References to a particular protein (*e.g.* '287', '919' *etc.*) may be taken to include that protein from any strain.

Non-fusion expression

In a first approach to heterologous expression, no fusion partner is used, and the native leader peptide (if present) is used. This will typically prevent any 'interference' from fusion partners and may alter cellular localisation and/or post-translational modification and/or folding in the heterologous host.

Thus the invention provides a method for the heterologous expression of a protein of the invention, in which (a) no fusion partner is used, and (b) the protein's native leader peptide (if present) is used.

The method will typically involve the step of preparing an vector for expressing a protein of the invention, such that the first expressed amino acid is the first amino acid (methionine) of said protein, and last expressed amino acid is the last amino acid of said protein (*i.e.* the codon preceding the native STOP codon).

This approach is preferably used for the expression of the following proteins using the native leader peptide: 111, 149, 206, 225-1, 235, 247-1, 274, 283, 286, 292, 401, 406, 502-1, 503, 519-1, 525-1, 552, 556, 557, 570, 576-1, 580, 583, 664, 759, 907, 913, 920-1, 936-1, 953, 961, 983, 989, Orf4, Orf7-1, Orf9-1, Orf23, Orf25, Orf37, Orf38, Orf40, Orf40.1, Orf40.2, Orf72-1, Orf76-1, Orf85-2, Orf91, Orf97-1, Orf119, Orf143.1, NMB0109 and NMB2050. The suffix 'L' used herein in the name of a protein indicates expression in this manner using the native leader peptide.

Proteins which are preferably expressed using this approach using no fusion partner and which have no native leader peptide include: 008, 105, 117-1, 121-1, 122-1, 128-1, 148, 216, 243, 308, 593, 652, 726, 926, 982, Orf83-1 and Orf143-1.

Advantageously, it is used for the expression of ORF25 or ORF40, resulting in a protein which induces better anti-bactericidal antibodies than GST- or His-fusions.

This approach is particularly suited for expressing lipoproteins.

Leader-peptide substitution

In a second approach to heterologous expression, the native leader peptide of a protein of the invention is replaced by that of a different protein. In addition, it is preferred that no fusion partner is used. Whilst using a protein's own leader peptide in heterologous hosts can often localise the protein to its 'natural' cellular location, in some cases the leader sequence is not efficiently recognised by the heterologous host. In such cases, a leader peptide known to drive protein targeting efficiently can be used instead.

Thus the invention provides a method for the heterologous expression of a protein of the invention, in which (a) the protein's leader peptide is replaced by the leader peptide from a different protein and, optionally, (b) no fusion partner is used.

The method will typically involve the steps of: obtaining nucleic acid encoding a protein of the invention; manipulating said nucleic acid to remove nucleotides that encode the protein's leader peptide and to introduce nucleotides that encode a different protein's leader peptide. The resulting nucleic acid may be inserted into an expression vector, or may already be part of an expression vector. The expressed protein will consist of the replacement leader peptide at the N-terminus, followed by the protein of the invention minus its leader peptide.

The leader peptide is preferably from another protein of the invention (*e.g.* one of SEQ#s 1-4326), but may also be from an *E.coli* protein (*e.g.* the OmpA leader peptide) or an *Erwinia carotovora* protein (*e.g.* the PelB leader peptide), for instance.

A particularly useful replacement leader peptide is that of ORF4. This leader is able to direct lipidation in *E.coli*, improving cellular localisation, and is particularly useful for the expression of proteins 287, 919 and ΔG287. The leader peptide and N-terminal domains of 961 are also particularly useful.

Another useful replacement leader peptide is that of *E.coli* OmpA. This leader is able to direct membrane localisation of *E.coli*. It is particularly advantageous for the expression of ORF1, resulting in a protein which induces better anti-bactericidal antibodies than both fusions and protein expressed from its own leader peptide.

Another useful replacement leader peptide is MKKYLFSAA. This can direct secretion into culture medium, and is extremely short and active. The use of this leader peptide is not restricted to the expression of Neisserial proteins – it may be used to direct the expression of any protein (particularly bacterial proteins).

Leader-peptide deletion

In a third approach to heterologous expression, the native leader peptide of a protein of the invention is deleted. In addition, it is preferred that no fusion partner is used.

Thus the invention provides a method for the heterologous expression of a protein of the invention, in which (a) the protein's leader peptide is deleted and, optionally, (b) no fusion partner is used.

The method will typically involve the steps of: obtaining nucleic acid encoding a protein of the invention; manipulating said nucleic acid to remove nucleotides that encode the protein's leader peptide. The resulting nucleic acid may be inserted into an expression vector, or may

already be part of an expression vector. The first amino acid of the expressed protein will be that of the mature native protein.

This method can increase the levels of expression. For protein 919, for example, expression levels in *E.coli* are much higher when the leader peptide is deleted. Increased expression may be due to altered localisation in the absence of the leader peptide.

The method is preferably used for the expression of 919, ORF46, 961, 050-1, 760 and 287.

Domain-based expression

In a fourth approach to heterologous expression, the protein is expressed as domains. This may be used in association with fusion systems (*e.g.* GST or His-tag fusions).

Thus the invention provides a method for the heterologous expression of a protein of the invention, in which (a) at least one domain in the protein is deleted and, optionally, (b) no fusion partner is used.

The method will typically involve the steps of: obtaining nucleic acid encoding a protein of the invention; manipulating said nucleic acid to remove at least one domain from within the protein. The resulting nucleic acid may be inserted into an expression vector, or may already be part of an expression vector. Where no fusion partners are used, the first amino acid of the expressed protein will be that of a domain of the protein.

A protein is typically divided into notional domains by aligning it with known sequences in databases and then determining regions of the protein which show different alignment patterns from each other.

The method is preferably used for the expression of protein 287. This protein can be notionally split into three domains, referred to as A B & C (see Figure 5). Domain B aligns strongly with IgA proteases, domain C aligns strongly with transferrin-binding proteins, and domain A shows no strong alignment with database sequences. An alignment of polymorphic forms of 287 is disclosed in WO00/66741.

Once a protein has been divided into domains, these can be (a) expressed singly (b) deleted from with the protein *e.g.* protein ABCD → ABD, ACD, BCD *etc.* or (c) rearranged *e.g.* protein ABC → ACB, CAB *etc.* These three strategies can be combined with fusion partners is desired.

ORF46 has also been notionally split into two domains – a first domain (amino acids 1-433) which is well-conserved between species and serogroups, and a second domain (amino acids 433-608) which is not well-conserved. The second domain is preferably deleted. An alignment of polymorphic forms of ORF46 is disclosed in WO00/66741.

Protein 564 has also been split into domains (Figure 8), as have protein 961 (Figure 12) and protein 502 (amino acids 28-167 of the MC58 protein).

Hybrid proteins

In a fifth approach to heterologous expression, two or more (*e.g.* 3, 4, 5, 6 or more) proteins of the invention are expressed as a single hybrid protein. It is preferred that no non-Neisserial fusion partner (*e.g.* GST or poly-His) is used.

This offers two advantages. Firstly, a protein that may be unstable or poorly expressed on its own can be assisted by adding a suitable hybrid partner that overcomes the problem. Secondly, commercial manufacture is simplified – only one expression and purification need be employed in order to produce two separately-useful proteins.

Thus the invention provides a method for the simultaneous heterologous expression of two or more proteins of the invention, in which said two or more proteins of the invention are fused (*i.e.* they are translated as a single polypeptide chain).

The method will typically involve the steps of: obtaining a first nucleic acid encoding a first protein of the invention; obtaining a second nucleic acid encoding a second protein of the invention; ligating the first and second nucleic acids. The resulting nucleic acid may be inserted into an expression vector, or may already be part of an expression vector.

Preferably, the constituent proteins in a hybrid protein according to the invention will be from the same strain.

The fused proteins in the hybrid may be joined directly, or may be joined via a linker peptide *e.g.* via a poly-glycine linker (*i.e.* G_n where n = 3, 4, 5, 6, 7, 8, 9, 10 or more) or via a short peptide sequence which facilitates cloning. It is evidently preferred not to join a ΔG protein to the C-terminus of a poly-glycine linker.

The fused proteins may lack native leader peptides or may include the leader peptide sequence of the N-terminal fusion partner.

The method is well suited to the expression of proteins orf1, orf4, orf25, orf40, Orf46/46.1, orf83, 233, 287, 292L, 564, 687, 741, 907, 919, 953, 961 and 983.

The 42 hybrids indicated by 'X' in the following table of form NH_2-A—B-COOH are preferred:

↓A B→	ORF46.1	287	741	919	953	961	983
ORF46.1		X	X	X	X	X	X
287	X		X	X	X	X	X
741	X	X		X	X	X	X
919	X	X	X		X	X	X
953	X	X	X	X		X	X
961	X	X	X	X	X		X
983	X	X	X	X	X	X	

Preferred proteins to be expressed as hybrids are thus ORF46.1, 287, 741, 919, 953, 961 and 983. These may be used in their essentially full-length form, or poly-glycine deletions (ΔG) forms may be used (*e.g.* ΔG-287, ΔGTbp2, ΔG741, ΔG983 *etc.*), or truncated forms may be used (*e.g.* Δ1-287, Δ2-287 *etc.*), or domain-deleted versions may be used (*e.g.* 287B, 287C, 287BC, $ORF46_{1-433}$, $ORF46_{433-608}$, ORF46, 961c *etc.*).

Particularly preferred are: (a) a hybrid protein comprising 919 and 287; (b) a hybrid protein comprising 953 and 287; (c) a hybrid protein comprising 287 and ORF46.1; (d) a hybrid protein comprising ORF1 and ORF46.1; (e) a hybrid protein comprising 919 and ORF46.1; (f) a hybrid protein comprising ORF46.1 and 919; (g) a hybrid protein comprising ORF46.1, 287 and 919; (h) a hybrid protein comprising 919 and 519; and (i) a hybrid protein comprising ORF97 and 225. Further embodiments are shown in Figure 14.

Where 287 is used, it is preferably at the C-terminal end of a hybrid; if it is to be used at the N-terminus, if is preferred to use a ΔG form of 287 is used (*e.g.* as the N-terminus of a hybrid with ORF46.1, 919, 953 or 961).

Where 287 is used, this is preferably from strain 2996 or from strain 394/98.

Where 961 is used, this is preferably at the N-terminus. Domain forms of 961 may be used.

Alignments of polymorphic forms of ORF46, 287, 919 and 953 are disclosed in WO00/66741. Any of these polymorphs can be used according to the present invention.

Temperature

In a sixth approach to heterologous expression, proteins of the invention are expressed at a low temperature.

Expressed Neisserial proteins (*e.g.* 919) may be toxic to *E.coli*, which can be avoided by expressing the toxic protein at a temperature at which its toxic activity is not manifested.

Thus the present invention provides a method for the heterologous expression of a protein of the invention, in which expression of a protein of the invention is carried out at a temperature at which a toxic activity of the protein is not manifested.

A preferred temperature is around 30°C. This is particularly suited to the expression of 919.

Mutations

As discussed above, expressed Neisserial proteins may be toxic to *E.coli*. This toxicity can be avoided by mutating the protein to reduce or eliminate the toxic activity. In particular, mutations to reduce or eliminate toxic enzymatic activity can be used, preferably using site-directed mutagenesis.

In a seventh approach to heterologous expression, therefore, an expressed protein is mutated to reduce or eliminate toxic activity.

Thus the invention provides a method for the heterologous expression of a protein of the invention, in which protein is mutated to reduce or eliminate toxic activity.

The method is preferably used for the expression of protein 907, 919 or 922. A preferred mutation in 907 is at Glu-117 (*e.g.* Glu→Gly); preferred mutations in 919 are at Glu-255 (*e.g.* Glu→Gly) and/or Glu-323 (*e.g.* Glu→Gly); preferred mutations in 922 are at Glu-164 (*e.g.* Glu→Gly), Ser-213 (*e.g.* Ser→Gly) and/or Asn-348 (*e.g.* Asn→Gly).

Alternative vectors

In a eighth approach to heterologous expression, an alternative vector used to express the protein. This may be to improve expression yields, for instance, or to utilise plasmids that are already approved for GMP use.

Thus the invention provides a method for the heterologous expression of a protein of the invention, in which an alternative vector is used. The alternative vector is preferably pSM214, with no fusion partners. Leader peptides may or may not be included.

This approach is particularly useful for protein 953. Expression and localisation of 953 with its native leader peptide expressed from pSM214 is much better than from the pET vector.

pSM214 may also be used with: ΔG287, Δ2-287, Δ3-287, Δ4-287, Orf46.1, 961L, 961, 961(MC58), 961c, 961c-L, 919, 953 and ΔG287-Orf46.1.

Another suitable vector is pET-24b (Novagen; uses kanamycin resistance), again using no fusion partners. pET-24b is preferred for use with: ΔG287K, Δ2-287K, Δ3-287K, Δ4-287K, Orf46.1-K, Orf46A-K, 961-K (MC58), 961a-K, 961b-K, 961c-K, 961c-L-K, 961d-K, ΔG287-919-K, ΔG287-Orf46.1-K and ΔG287-961-K.

Multimeric form

In a ninth approach to heterologous expression, a protein is expressed or purified such that it adopts a particular multimeric form.

This approach is particularly suited to protein 953. Purification of one particular multimeric form of 953 (the monomeric form) gives a protein with greater bactericidal activity than other forms (the dimeric form).

Proteins 287 and 919 may be purified in dimeric forms.

Protein 961 may be purified in a 180kDa oligomeric form (*e.g.* a tetramer).

Lipidation

In a tenth approach to heterologous expression, a protein is expressed as a lipidated protein.

Thus the invention provides a method for the heterologous expression of a protein of the invention, in which the protein is expressed as a lipidated protein.

This is particularly useful for the expression of 919, 287, ORF4, 406, 576-1, and ORF25. Polymorphic forms of 919, 287 and ORF4 are disclosed in WO00/66741.

The method will typically involve the use of an appropriate leader peptide without using an N-terminal fusion partner.

C-terminal deletions

In an eleventh approach to heterologous expression, the C-terminus of a protein of the invention is mutated. In addition, it is preferred that no fusion partner is used.

Thus the invention provides a method for the heterologous expression of a protein of the invention, in which (a) the protein's C-terminus region is mutated and, optionally, (b) no fusion partner is used.

The method will typically involve the steps of: obtaining nucleic acid encoding a protein of the invention; manipulating said nucleic acid to mutate nucleotides that encode the protein's C-terminus portion. The resulting nucleic acid may be inserted into an expression vector, or may already be part of an expression vector. The first amino acid of the expressed protein will be that of the mature native protein.

The mutation may be a substitution, insertion or, preferably, a deletion.

This method can increase the levels of expression, particularly for proteins 730, ORF29 and ORF46. For protein 730, a C-terminus region of around 65 to around 214 amino acids may be deleted; for ORF46, the C-terminus region of around 175 amino acids may be deleted; for ORF29, the C-terminus may be deleted to leave around 230-370 N-terminal amino acids.

Leader peptide mutation

In a twelfth approach to heterologous expression, the leader peptide of the protein is mutated. This is particularly useful for the expression of protein 919.

Thus the invention provides a method for the heterologous expression of a protein of the invention, in which the protein's leader peptide is mutated.

The method will typically involve the steps of: obtaining nucleic acid encoding a protein of the invention; and manipulating said nucleic acid to mutate nucleotides within the leader peptide. The resulting nucleic acid may be inserted into an expression vector, or may already be part of an expression vector.

Poly-glycine deletion

In a thirteenth approach to heterologous expression, poly-glycine stretches in wild-type sequences are mutated. This enhances protein expression.

The poly-glycine stretch has the sequence $(Gly)_n$, where $n \geq 4$ (*e.g.* 5, 6, 7, 8, 9 or more). This stretch is mutated to disrupt or remove the $(Gly)_n$. This may be by deletion (*e.g.* CGGGGS→ CGGGS, CGGS, CGS or CS), by substitution (*e.g.* CGGGGS→ CGXGGS, CGXXGS, CGXGXS *etc.*), and/or by insertion (*e.g.* CGGGGS→ CGGXGGS, CGXGGGS, *etc.*).

This approach is not restricted to Neisserial proteins – it may be used for any protein (particularly bacterial proteins) to enhance heterologous expression. For Neisserial proteins, however, it is particularly suitable for expressing 287, 741, 983 and Tbp2. An alignment of polymorphic forms of 287 is disclosed in WO00/66741.

Thus the invention provides a method for the heterologous expression of a protein of the invention, in which (a) a poly-glycine stretch within the protein is mutated.

The method will typically involve the steps of: obtaining nucleic acid encoding a protein of the invention; and manipulating said nucleic acid to mutate nucleotides that encode a poly-glycine stretch within the protein sequence. The resulting nucleic acid may be inserted into an expression vector, or may already be part of an expression vector.

Conversely, the opposite approach (*i.e.* introduction of poly-glycine stretches) can be used to suppress or diminish expression of a given heterologous protein.

Heterologous host

Whilst expression of the proteins of the invention may take place in the native host (*i.e.* the organism in which the protein is expressed in nature), the present invention utilises a heterologous host. The heterologous host may be prokaryotic or eukaryotic. It is preferably *E.coli*, but other suitable hosts include *Bacillus subtilis, Vibrio cholerae, Salmonella typhi, Salmonenna typhimurium, Neisseria meningitidis, Neisseria gonorrhoeae, Neisseria lactamica, Neisseria cinerea, Mycobateria* (*e.g. M.tuberculosis*), yeast *etc.*

***Vectors* etc.**

As well as the methods described above, the invention provides (a) nucleic acid and vectors useful in these methods (b) host cells containing said vectors (c) proteins expressed or expressable by the methods (d) compositions comprising these proteins, which may be suitable as vaccines, for instance, or as diagnostic reagents, or as immunogenic compositions (e) these compositions for use as medicaments (*e.g.* as vaccines) or as diagnostic reagents (f) the use of these compositions in the manufacture of (1) a medicament for treating or preventing infection due to Neisserial bacteria (2) a diagnostic reagent for detecting the presence of Neisserial bacteria or of antibodies raised against Neisserial bacteria, and/or (3) a reagent which can raise antibodies against Neisserial bacteria and (g) a method of treating a

patient, comprising administering to the patient a therapeutically effective amount of these compositions.

Sequences

The invention also provides a protein or a nucleic acid having any of the sequences set out in the following examples. It also provides proteins and nucleic acid having sequence identity to these. As described above, the degree of 'sequence identity' is preferably greater than 50% (eg. 60%, 70%, 80%, 90%, 95%, 99% or more).

Furthermore, the invention provides nucleic acid which can hybridise to the nucleic acid disclosed in the examples, preferably under "high stringency" conditions (*eg.* 65°C in a 0.1xSSC, 0.5% SDS solution).

The invention also provides nucleic acid encoding proteins according to the invention.

It should also be appreciated that the invention provides nucleic acid comprising sequences complementary to those described above (*eg.* for antisense or probing purposes).

Nucleic acid according to the invention can, of course, be prepared in many ways (*eg.* by chemical synthesis, from genomic or cDNA libraries, from the organism itself *etc.*) and can take various forms (*eg.* single stranded, double stranded, vectors, probes *etc.*).

In addition, the term "nucleic acid" includes DNA and RNA, and also their analogues, such as those containing modified backbones, and also peptide nucleic acids (PNA) *etc.*

BRIEF DESCRIPTION OF DRAWINGS

Figures 1 and 2 show constructs used to express proteins using heterologous leader peptides.

Figure 3 shows expression data for ORF1, and Figure 4 shows similar data for protein 961.

Figure 5 shows domains of protein 287, and Figures 6 & 7 show deletions within domain A.

Figure 8 shows domains of protein 564.

Figure 9 shows the *PhoC* reporter gene driven by the 919 leader peptide, and Figure 10 shows the results obtained using mutants of the leader peptide.

Figure 11 shows insertion mutants of protein 730 (A: 730-C1; B: 730-C2).

Figure 12 shows domains of protein 961.

Figure 13 shows SDS-PAGE of ΔG proteins. Dots show the main recombinant product.

Figure 14 shows 26 hybrid proteins according to the invention.

MODES FOR CARRYING OUT THE INVENTION

Example 1 – 919 and its leader peptide

Protein 919 from *N.meningitidis* (serogroup B, strain 2996) has the following sequence:

```
  1  MKKYLFRAAL YGIAAAILAA CQSKSIQTFP QPDTSVINGP DRPVGIPDPA
 51  GTTVGGGGAV YTVVPHLSLP HWAAQDFAKS LQSFRLGCAN LKNRQGWQDV
101  CAQAFQTPVH SFQAKQFFER YFTPWQVAGN GSLAGTVTGY YEPVLKGDDR
151  RTAQARFPIY GIPDDFISVP LPAGLRSGKA LVRIRQTGKN SGTIDNTGGT
201  HTADLSRFPI TARTTAIKGR FEGSRFLPYH TRNQINGGAL DGKAPILGYA
251  EDPVELFFMH IQGSGRLKTP SGKYIRIGYA DKNEHPYVSI GRYMADKGYL
301  KLGQTSMQGI KAYMRQNPQR LAEVLGQNPS YIFFRELAGS SNDGPVGALG
351  TPLMGEYAGA VDRHYITLGA PLFVATAHPV TRKALNRLIM AQDTGSAIKG
401  AVRVDYFWGY GDEAGELAGK QKTTGYVWQL LPNGMKPEYR P*
```

The leader peptide is underlined.

The sequences of 919 from other strains can be found in Figures 7 and 18 of WO00/66741.

Example 2 of WO99/57280 discloses the expression of protein 919 as a His-fusion in *E.coli*. The protein is a good surface-exposed immunogen.

Three alternative expression strategies were used for 919:

1) 919 without its leader peptide (and without the mature N-terminal cysteine) and without any fusion partner ('$919^{untagged}$'):

```
  1  QSKSIQTFP QPDTSVINGP DRPVGIPDPA GTTVGGGGAV YTVVPHLSLP
 50  HWAAQDFAKS LQSFRLGCAN LKNRQGWQDV CAQAFQTPVH SFQAKQFFER
100  YFTPWQVAGN GSLAGTVTGY YEPVLKGDDR RTAQARFPIY GIPDDFISVP
150  LPAGLRSGKA LVRIRQTGKN SGTIDNTGGT HTADLSRFPI TARTTAIKGR
200  FEGSRFLPYH TRNQINGGAL DGKAPILGYA EDPVELFFMH IQGSGRLKTP
250  SGKYIRIGYA DKNEHPYVSI GRYMADKGYL KLGQTSMQGI KAYMRQNPQR
300  LAEVLGQNPS YIFFRELAGS SNDGPVGALG TPLMGEYAGA VDRHYITLGA
350  PLFVATAHPV TRKALNRLIM AQDTGSAIKG AVRVDYFWGY GDEAGELAGK
400  QKTTGYVWQL LPNGMKPEYR P*
```

The leader peptide and cysteine were omitted by designing the 5'-end amplification primer downstream from the predicted leader sequence.

2) 919 with its own leader peptide but without any fusion partner ('919L'); and

3) 919 with the leader peptide (MKTFFKTLSAAALALILAA) from ORF4 ('919LOrf4').

```
  1  MKTFFKTLS AAALALILAA CQSKSIQTFP QPDTSVINGP DRPVGIPDPA
 50  GTTVGGGGAV YTVVPHLSLP HWAAQDFAKS LQSFRLGCAN LKNRQGWQDV
100  CAQAFQTPVH SFQAKQFFER YFTPWQVAGN GSLAGTVTGY YEPVLKGDDR
150  RTAQARFPIY GIPDDFISVP LPAGLRSGKA LVRIRQTGKN SGTIDNTGGT
200  HTADLSRFPI TARTTAIKGR FEGSRFLPYH TRNQINGGAL DGKAPILGYA
250  EDPVELFFMH IQGSGRLKTP SGKYIRIGYA DKNEHPYVSI GRYMADKGYL
300  KLGQTSMQGI KSYMRQNPQR LAEVLGQNPS YIFFRELAGS SNDGPVGALG
```

350 TPLMGEYAGA VDRHYITLGA PLFVATAHPV TRKALNRLIM AQDTGSAIKG
400 AVRVDYFWGY GDEAGELAGK QKTTGYVWQL LPNGMKPEYR P*

To make this construct, the entire sequence encoding the ORF4 leader peptide was included in the 5'-primer as a tail (primer 919Lorf4 For). A *Nhe*I restriction site was generated by a double nucleotide change in the sequence coding for the ORF4 leader (no amino acid changes), to allow different genes to be fused to the ORF4 leader peptide sequence. A stop codon was included in all the 3'-end primer sequences.

All three forms of the protein were expressed and could be purified.

The '919L' and '919LOrf4' expression products were both lipidated, as shown by the incorporation of [^{3}H]-palmitate label. $919^{untagged}$ did not incorporate the ^{3}H label and was located intracellularly.

919LOrf4 could be purified more easily than 919L. It was purified and used to immunise mice. The resulting sera gave excellent results in FACS and ELISA tests, and also in the bactericidal assay. The lipoprotein was shown to be localised in the outer membrane.

$919^{untagged}$ gave excellent ELISA titres and high serum bactericidal activity. FACS confirmed its cell surface location.

Example 2 – 919 and expression temperature

Growth of *E.coli* expressing the 919LOrf4 protein at 37°C resulted in lysis of the bacteria. In order to overcome this problem, the recombinant bacteria were grown at 30°C. Lysis was prevented without preventing expression.

Example 3 – mutation of 907, 919 and 922

It was hypothesised that proteins 907, 919 and 922 are murein hydrolases, and more particularly lytic transglycosylases. Murein hydrolases are located on the outer membrane and participate in the degradation of peptidoglycan.

The purified proteins $919^{untagged}$, 919Lorf4, 919-His (*i.e.* with a C-terminus His-tag) and 922-His were thus tested for murein hydrolase activity [Ursinus & Holtje (1994) *J.Bact.* 176:338-343]. Two different assays were used, one determining the degradation of insoluble murein sacculus into soluble muropeptides and the other measuring breakdown of poly(MurNAc-GlcNAc)$_{n>30}$ glycan strands.

The first assay uses murein sacculi radiolabelled with meso-2,6-diamino-3,4,5-[^{3}H]pimelic acid as substrate. Enzyme (3–10 µg total) was incubated for 45 minutes at 37°C in a total volume of 100µl comprising 10mM Tris-maleate (pH 5.5), 10mM $MgCl_2$, 0.2% v/v Triton X-100 and [^{3}H]A_2pm labelled murein sacculi (about 10000cpm). The assay mixture was placed on ice for 15 minutes with 100 µl of 1% w/v N-acetyl-N,N,N-trimethylammonium for 15 minutes and precipitated material pelleted by centrifugation at 10000*g* for 15 minutes. The radioactivity in the supernatant was measured by liquid scintillation counting. *E.coli* soluble lytic transglycosylase Slt70 was used as a positive control for the assay; the negative control comprised the above assay solution without enzyme.

All proteins except 919-His gave positive results in the first assay.

The second assay monitors the hydrolysis of poly(MurNAc-GlcNAc)glycan strands. Purified strands, poly(MurNAc-GlcNAc)$_{n>30}$ labelled with N-acetyl-D-1-[^{3}H]glucosamine were incubated with 3µg of 919L in 10 mM Tris-maleate (pH 5.5), 10 mM $MgCl_2$ and 0.2% v/v Triton X-100 for 30 min at 37°C. The reaction was stopped by boiling for 5 minutes and the pH of the sample adjusted to about 3.5 by addition of 10µl of 20% v/v phosphoric acid. Substrate and product were separated by reversed phase HPLC on a Nucleosil 300 C_{18} column as described by Harz *et. al.* [*Anal. Biochem.* (1990) 190:120-128]. The *E.coli* lytic transglycosylase Mlt A was used as a positive control in the assay. The negative control was performed in the absence of enzyme.

By this assay, the ability of 919LOrf4 to hydrolyse isolated glycan strands was demonstrated when anhydrodisaccharide subunits were separated from the oligosaccharide by HPLC.

Protein 919Lorf4 was chosen for kinetic analyses. The activity of 919Lorf4 was enhanced 3.7-fold by the addition of 0.2% v/v Triton X-100 in the assay buffer. The presence of Triton X-100 had no effect on the activity of $919^{untagged}$. The effect of pH on enzyme activity was determined in Tris-Maleate buffer over a range of 5.0 to 8.0. The optimal pH for the reaction was determined to be 5.5. Over the temperature range 18°C to 42°C, maximum activity was observed at 37°C. The effect of various ions on murein hydrolase activity was determined by performing the reaction in the presence of a variety of ions at a final concentration of 10mM. Maximum activity was found with Mg^{2+}, which stimulated activity 2.1-fold. Mn^{2+} and Ca^{2+} also stimulated enzyme activity to a similar extent while the addition Ni^{2+} and EDTA had no significant effect. In contrast, both $Fe^{2+}$and Zn^{2+} significantly inhibited enzyme activity.

The structures of the reaction products resulting from the digestion of unlabelled *E.coli* murein sacculus were analysed by reversed-phase HPLC as described by Glauner [*Anal. Biochem.* (1988) 172:451-464]. Murein sacculi digested with the muramidase Cellosyl were used to calibrate and standardise the Hypersil ODS column. The major reaction products were 1,6 anhydrodisaccharide tetra and tri peptides, demonstrating the formation of 1,6 anhydromuraminic acid intramolecular bond.

These results demonstrate experimentally that 919 is a murein hydrolase and in particular a member of the lytic transglycosylase family of enzymes. Furthermore the ability of 922-His to hydrolyse murein sacculi suggests this protein is also a lytic transglycosylase.

This activity may help to explain the toxic effects of 919 when expressed in *E.coli*.

In order to eliminate the enzymatic activity, rational mutagenesis was used. 907, 919 and 922 show fairly low homology to three membrane-bound lipidated murein lytic transglycosylases from *E.coli*:

919 (441aa) is 27.3% identical over 440aa overlap to *E.coli* MLTA (P46885);

922 (369aa) is 38.7% identical over 310aa overlap to *E.coli* MLTB (P41052); and

907-2 (207aa) is 26.8% identical over 149aa overlap to *E.coli* MLTC (P52066).

907-2 also shares homology with *E.coli* MLTD (P23931) and Slt70 (P03810), a soluble lytic transglycosylase that is located in the periplasmic space. No significant sequence homology can be detected among 919, 922 and 907-2, and the same is true among the corresponding MLTA, MLTB and MLTC proteins.

Crystal structures are available for Slt70 [1QTEA; 1QTEB; Thunnissen *et al.* (1995) *Biochemistry* 34:12729-12737] and for Slt35 [1LTM; 1QUS; 1QUT; van Asselt *et al.* (1999) *Structure Fold Des* 7:1167-80] which is a soluble form of the 40kDa MLTB.

The catalytic residue (a glutamic acid) has been identified for both Slt70 and MLTB.

In the case of Slt70, mutagenesis studies have demonstrated that even a conservative substitution of the catalytic Glu505 with a glutamine (Gln) causes the complete loss of enzymatic activity. Although Slt35 has no obvious sequence similarity to Slt70, their catalytic domains shows a surprising similarity. The corresponding catalytic residue in MLTB is Glu162.

Another residue which is believed to play an important role in the correct folding of the enzymatic cleft is a well-conserved glycine (Gly) downstream of the glutamic acid. Recently, Terrak *et al.* [*Mol.Microbiol.* (1999) 34:350-64] have suggested the presence of another important residue which is an aromatic amino acid located around 70-75 residues downstream of the catalytic glutamic acid.

Sequence alignment of Slt70 with 907-2 and of MLTB with 922 were performed in order to identify the corresponding catalytic residues in the MenB antigens.

The two alignments in the region of the catalytic domain are reported below:

907-2/Slt70:

```
                 90        100         110        ▼120        130        140
907-2.pep        ERRRLLVNIQYESSRAG--LDTQIVLGLIEVESAFRQYAISGVGARGLMQVMPFWKNYIG
                 ||  |  |  ::   :|  :  : :::: : |||:   : | ||| ||||:||   ::
slty_ecoli       ERFPLAYNDLFKRYTSGKEIPQSYAMAIARQESAWNPKVKSPVGASGLMQIMPGTATHTV
                     480        490        500     ▲ 510        520        530
                                                   GLU505
```

922/MLTB

```
                   150        160   ▼    170        180        190        200
922.pep            VAQKYGVPAELIVAVIGIETNYGKNTGSFRVADALATLGFDYPRRAGFFQKELVELLKLA
                   : | |||| |:||::||:|| :|:  |: |: |||||||:|:||||| :|: ||  :| :|
mltb_ecoli         AWQVYGVPPEIIVGIIGVETRWGRVMGKTRILDALATLSFNYPRRAEYFSGELETFLLMA
                       150        160 ▲      170        180        190        200
                                      GLU162

                   210        220        230        240        250        260
922.pep            KEEGGDVFAFKGSYAGAMGMPQFMPSSYRKWAVDYDGDGHRDIWGNVGDVAASVANYMKQ
                   ::|  | : :|||:|||||: |||||||:::|||::|||| ::|  | |: :|||||:|
mltb_ecoli         RDEQDDPLNLKGSFAGAMGYGQFMPSSYKQYAVDFSGDGHINLWDPV-DAIGSVANYFKA
                       210        220        230        240        250        260
```

From these alignments, it results that the corresponding catalytic glutamate in 907-2 is Glu117, whereas in 922 is Glu164. Both antigens also share downstream glycines that could have a structural role in the folding of the enzymatic cleft (in bold), and 922 has a conserved aromatic residue around 70aa downstream (in bold).

In the case of protein 919, no 3D structure is available for its *E.coli* homologue MLTA, and nothing is known about a possible catalytic residue. Nevertheless, three amino acids in 919 are predicted as catalytic residues by alignment with MLTA:

919/MLTA

```
                240        250     ▼  260  □ □  270 □      280        290
919.pep          ALDGKAPILGYAEDPVELFFMHIQGSGRLKTPSGKYIRI-GYADKNEHPYVSIGRYMADK
                 ||: |  ||:|::: :: |:| :|||| :   :|: : : :|| || | | |||: : |:
mlta_ecoli.p     ALSDKY-ILAYSNSLMDNFIMDVQGSGYIDFGDGSPLNFFSYAGKNGHAYRSIGKVLIDR
                           170        180        190        200        210
```

```
                300       310       320  ▼    330□ □□   340       ◊350    ◊
919.pep         GYLKLGQTSMQGIKSYMRQNPQ-RLAEVLGQNPSYIFFRELAGSSNDGPV-GALGTPLMG
                | :|  : |||:|: : : : : :: |:| ||||::||:  : :    || || ::||:|
mlta_ecoli.p    GEVKKEDMSMQAIRHWGETHSEAEVRELLEQNPSFVFFKPQSFA----PVKGASAVPLVG
              220       230       240       250       260           270

                  360 ▼      o        380            390       400     ◊◊410
919.pep         EYAGAVDRHYITLGAPLFVATAHPVTRKALN-----RLIMAQDTGSAIKGAVRVDYFWGY
                : : | ||  |  |: |:: :    :   :|     ||::| |:|:|||| : | : |
mlta_ecoli.p    RASVASDRSIIPPGTTLLAEVPLLDNNGKFNGQYELRLMVALDVGGAIKGQ-HFDIYQGI
                  280       290       300       310       320       330

                       420      o
919.pep         GDEAGELAGKQKTTGYVWQLLP
                | |||: ||  :  | || |
mlta_ecoli.p    GPEAGHRAGWYNHYGRVWVLKT
                   340       350
```

The three possible catalytic residues are shown by the symbol ▼:

1) Glu255 (Asp in MLTA), followed by three conserved glycines (Gly263, Gly265 and Gly272) and three conserved aromatic residues located approximately 75-77 residues downstream. These downstream residues are shown by □.

2) Glu323 (conserved in MLTA), followed by 2 conserved glycines (Gly347 and Gly355) and two conserved aromatic residues located 84-85 residues downstream (Tyr406 or Phe407). These downstream residues are shown by ◊.

3) Asp362 (instead of the expected Glu), followed by one glycine (Gly 369) and a conserved aromatic residue (Trp428). These downstream residues are shown by o.

Alignments of polymorphic forms of 919 are disclosed in WO00/66741.

Based on the prediction of catalytic residues, three mutants of the 919 and one mutant of 907, containing each a single amino acid substitution, have been generated. The glutamic acids in position 255 and 323 and the aspartic acids in position 362 of the 919 protein and the glutamic acid in position 117 of the 907 protein, were replaced with glycine residues using PCR-based SDM. To do this, internal primers containing a codon change from Glu or Asp to Gly were designed:

Primers	Sequences	Codon change
919-E255 for	CGAAGACCCCGTCGgtCTTTTTTTTATG	GAA → Ggt
919-E255 rev	GTGCATAAAAAAAAGacCGACGGGGTCT	
919-E323 for	AACGCCTCGCCGgtGTTTTGGGTCA	GAA → Ggt
919-E323 rev	TTTGACCCAAAACacCGGCGAGGCG	
919-D362 for	TGCCGGCGCAGTCGgtCGGCACTACA	GAC → Ggt
919-D362 rev	TAATGTAGTGCCGacCGACTGCGCCG	
907-E117 for	TGATTGAGGTGGgtAGCGCGTTCCG	GAA → Ggt
907-E117 rev	GGCGGAACGCGCTacCCACCTCAAT	

Underlined nucleotides code for glycine; the mutated nucleotides are in lower case.

To generate the 919-E255, 919-E323 and 919-E362 mutants, PCR was performed using 20ng of the pET 919-LOrf4 DNA as template, and the following primer pairs:

1) Orf4L for / 919-E255 rev
2) 919-E255 for / 919L rev
3) Orf4L for / 919-E323 rev
4) 919-E323 for / 919L rev
5) Orf4L for / 919-D362 rev
6) 919-D362 for / 919L rev

The second round of PCR was performed using the product of PCR 1-2, 3-4 or 5-6 as template, and as forward and reverse primers the "Orf4L for" and "919L rev" respectively.

For the mutant 907-E117, PCR have been performed using 200ng of chromosomal DNA of the 2996 strain as template and the following primer pairs:

7) 907L for / 907-E117 rev
8) 907-E117 for / 907L rev

The second round of PCR was performed using the products of PCR 7 and 8 as templates and the oligos "907L for" and "907L rev" as primers.

The PCR fragments containing each mutation were processed following the standard procedure, digested with *Nde*I and *Xho*I restriction enzymes and cloned into pET-21b+ vector. The presence of each mutation was confirmed by sequence analysis.

Mutation of Glu117 to Gly in 907 is carried out similarly, as is mutation of residues Glu164, Ser213 and Asn348 in 922.

The E255G mutant of 919 shows a 50% reduction in activity; the E323G mutant shows a 70% reduction in activity; the E362G mutant shows no reduction in activity.

Example 4 – multimeric form

287-GST, $919^{untagged}$ and 953-His were subjected to gel filtration for analysis of quaternary structure or preparative purposes. The molecular weight of the native proteins was estimated using either FPLC Superose 12 (H/R 10/30) or Superdex 75 gel filtration columns (Pharmacia). The buffers used for chromatography for 287, 919 and 953 were 50 mM Tris-HCl (pH 8.0), 20 mM Bicine (pH 8.5) and 50 mM Bicine (pH 8.0), respectively.

Additionally each buffer contained 150-200 mM NaCl and 10% v/v glycerol. Proteins were dialysed against the appropriate buffer and applied in a volume of 200µl. Gel filtration was performed with a flow rate of 0.5 – 2.0 ml/min and the eluate monitored at 280nm. Fractions were collected and analysed by SDS-PAGE. Blue dextran 2000 and the molecular weight standards ribonuclease A, chymotrypsin A ovalbumin, albumin (Pharmacia) were used to calibrate the column. The molecular weight of the sample was estimated from a calibration curve of K_{av} *vs.* log M_r of the standards. Before gel filtration, 287-GST was digested with thrombin to cleave the GST moiety.

The estimated molecular weights for 287, 919 and 953-His were 73 kDa, 47 kDa and 43 kDa respectively. These results suggest 919 is monomeric while both 287 and 953 are principally dimeric in their nature. In the case of 953-His, two peaks were observed during gel filtration. The major peak (80%) represented a dimeric conformation of 953 while the minor peak (20%) had the expected size of a monomer. The monomeric form of 953 was found to have greater bactericidal activity than the dimer.

Example 5 – pSM214 and pET-24b vectors

953 protein with its native leader peptide and no fusion partners was expressed from the pET vector and also from pSM214 [Velati Bellini *et al.* (1991) *J. Biotechnol.* 18, 177-192].

The 953 sequence was cloned as a full-length gene into pSM214 using the *E. coli* MM294-1 strain as a host. To do this, the entire DNA sequence of the 953 gene (from ATG to the STOP codon) was amplified by PCR using the following primers:

953L for/2	CCGGAATTCTTATGAAAAAAATCATCTTCGCCGC	Eco RI
953L rev/2	GCCCAAGCTTTTATTGTTTGGCTGCCTCGATT	Hind III

which contain *Eco*RI and *Hind*III restriction sites, respectively. The amplified fragment was digested with *Eco*RI and *Hind*III and ligated with the pSM214 vector digested with the same two enzymes. The ligated plasmid was transformed into *E.coli* MM294-1 cells (by incubation in ice for 65 minutes at 37° C) and bacterial cells plated on LB agar containing 20μg/ml of chloramphenicol.

Recombinant colonies were grown over-night at 37°C in 4 ml of LB broth containing 20 μg/ml of chloramphenicol; bacterial cells were centrifuged and plasmid DNA extracted as and analysed by restriction with *Eco*RI and *Hind*III. To analyse the ability of the recombinant colonies to express the protein, they were inoculated in LB broth containing 20μg/ml of chloramphenicol and let to grown for 16 hours at 37°C. Bacterial cells were centrifuged and resuspended in PBS. Expression of the protein was analysed by SDS-PAGE and Coomassie Blue staining.

Expression levels were unexpectedly high from the pSM214 plasmid.

Oligos used to clone sequences into pSM-214 vectors were as follows:

ΔG287	Fwd	CCGGAATTCTTATG-TCGCCCGATGTTAAATCGGCGGA	EcoRI
(pSM-214)	Rev	GCCCAAGCTT-TCAATCCTGCTCTTTTTTGCCG	HindIII
Δ2 287	Fwd	CCGGAATTCTTATG-AGCCAAGATATGGCGGCAGT	EcoRI
(pSM-214)	Rev	GCCCAAGCTT-TCAATCCTGCTCTTTTTTGCCG	HindIII
Δ3 287	Fwd	CCGGAATTCTTATG-TCCGCCGAATCCGCAAATCA	EcoRI
(pSM-214)	Rev	GCCCAAGCTT-TCAATCCTGCTCTTTTTTGCCG	HindIII
Δ4 287	Fwd	CCGGAATTCTTATG-GGAAGGGTTGATTTGGCTAATG	EcoRI
(pSM-214)	Rev	GCCCAAGCTT-TCAATCCTGCTCTTTTTTGCCG	HindIII
Orf46.1	Fwd	CCGGAATTCTTATG-TCAGATTTGGCAAACGATTCTT	EcoRI
(pSM-214)	Rev	GCCCAAGCTT-TTACGTATCATATTTCACGTGCTTC	HindIII
ΔG287-Orf46.1	Fwd	CCGGAATTCTTATG-TCGCCCGATGTTAAATCGGCGGA	EcoRI
(pSM-214)	Rev	GCCCAAGCTT-TTACGTATCATATTTCACGTGCTTC	HindIII
919	Fwd	CCGGAATTCTTATG-CAAAGCAAGAGCATCCAAACCT	EcoRI
(pSM-214)	Rev	GCCCAAGCTT-TTACGGGCGGTATTCGGGCT	HindIII
961L	Fwd	CCGGAATTCATATG-AAACACTTTCCATCC	EcoRI
(pSM-214)	Rev	GCCCAAGCTT-TTACCACTCGTAATTGAC	HindIII
961	Fwd	CCGGAATTCATATG-GCCACAAGCGACGAC	EcoRI
(pSM-214)	Rev	GCCCAAGCTT-TTACCACTCGTAATTGAC	HindIII
961c L	Fwd	CCGGAATTCTTATG-AAACACTTTCCATCC	EcoRI
pSM-214	Rev	GCCCAAGCTT-TCAACCCACGTTGTAAGGTTG	HindIII
961c	Fwd	CCGGAATTCTTATG-GCCACAAACGACGACG	EcoRI
pSM-214	Rev	GCCCAAGCTT-TCAACCCACGTTGTAAGGTTG	HindIII
953	Fwd	CCGGAATTCTTATG-GCCACCTACAAAGTGGACGA	EcoRI
(pSM-214)	Rev	GCCCAAGCTT-TTATTGTTTGGCTGCCTCGATT	HindIII

These sequences were manipulated, cloned and expressed as described for 953L.

For the pET-24 vector, sequences were cloned and the proteins expressed in pET-24 as described below for pET21. pET2 has the same sequence as pET-21, but with the kanamycin resistance cassette instead of ampicillin cassette.

Oligonucleotides used to clone sequences into pET-24b vector were:

ΔG 287 K	Fwd	CGCGGATCCGCTAGC-CCCGATGTTAAATCGGC [§]	NheI
	Rev	CCCGCTCGAG-TCAATCCTGCTCTTTTTTGCC *	XhoI
Δ2 287 K	Fwd	CGCGGATCCGCTAGC-CAAGATATGGCGGCAGT [§]	NheI
Δ3 287 K	Fwd	CGCGGATCCGCTAGC-GCCGAATCCGCAAATCA [§]	NheI
Δ4 287 K	Fwd	CGCGCTAGC-GGAAGGGTTGATTTGGCTAATGG [§]	NheI
Orf46.1 K	Fwd	GGGAATTCCATATG-GGCATTTCCCGCAAAATATC	NdeI
	Rev	CCCGCTCGAG-TTACGTATCATATTTCACGTGC	XhoI
Orf46A K	Fwd	GGGAATTCCATATG-GGCATTTCCCGCAAAATATC	NdeI
	Rev	CCCGCTCGAG-TTATTCTATGCCTTGTGCGGCAT	XhoI
961 K (MC58)	Fwd	CGCGGATCCCATATG-GCCACAAGCGACGACGA	NdeI
	Rev	CCCGCTCGAG-TTACCACTCGTAATTGAC	XhoI
961a K	Fwd	CGCGGATCCCATATG-GCCACAAACGACG	NdeI
	Rev	CCCGCTCGAG-TCATTTAGCAATATTATCTTTGTTC	XhoI
961b K	Fwd	CGCGGATCCCATATG-AAAGCAAACAGTGCCGAC	NdeI
	Rev	CCCGCTCGAG-TTACCACTCGTAATTGAC	XhoI
961c K	Fwd	CGCGGATCCCATATG-GCCACAAACGACG	NdeI
	Rev	CCCGCTCGAG-TTAACCCACGTTGTAAGGT	XhoI
961cL K	Fwd	CGCGGATCCCATATG-ATGAAACACTTTCCATCC	NdeI
	Rev	CCCGCTCGAG-TTAACCCACGTTGTAAGGT	XhoI
961d K	Fwd	CGCGGATCCCATATG-GCCACAAACGACG	NdeI
	Rev	CCCGCTCGAG-TCAGTCTGACACTGTTTTATCC	XhoI
ΔG 287-919 K	Fwd	CGCGGATCCGCTAGC-CCCGATGTTAAATCGGC	NheI
	Rev	CCCGCTCGAG-TTACGGGCGGTATTCGG	XhoI
ΔG 287-Orf46.1 K	Fwd	CGCGGATCCGCTAGC-CCCGATGTTAAATCGGC	NheI
	Rev	CCCGCTCGAG-TTACGTATCATATTTCACGTGC	XhoI
ΔG 287-961 K	Fwd	CGCGGATCCGCTAGC-CCCGATGTTAAATCGGC	NheI
	Rev	CCCGCTCGAG-TTACCACTCGTAATTGAC	XhoI

* This primer was used as a Reverse primer for all the 287 forms.

[§] Forward primers used in combination with the ΔG278 K reverse primer.

Example 6 – ORF1 and its leader peptide

ORF1 from *N.meningitidis* (serogroup B, strain MC58) is predicted to be an outer membrane or secreted protein. It has the following sequence:

1 MKTTDKRTTE THRKAPKTGR IRFSPAYLAI CLSFGILPQA WAGHTYFGIN

```
  51  YQYYRDFAEN KGKFAVGAKD IEVYNKKGEL VGKSMTKAPM IDFSVVSRNG
 101  VAALVGDQYI VSVAHNGGYN NVDFGAEGRN PDQHRFTYKI VKRNNYKAGT
 151  KGHPYGGDYH MPRLHKFVTD AEPVEMTSYM DGRKYIDQNN YPDRVRIGAG
 201  RQYWRSDEDE PNNRESSYHI ASAYSWLVGG NTFAQNGSGG GTVNLGSEKI
 251  KHSPYGFLPT GGSFGDSGSP MFIYDAQKQK WLINGVLQTG NPYIGKSNGF
 301  QLVRKDWFYD EIFAGDTHSV FYEPRQNGKY SFNDDNNGTG KINAKHEHNS
 351  LPNRLKTRTV QLFNVSLSET AREPVYHAAG GVNSYRPRLN NGENISFIDE
 401  GKGELILTSN INQGAGGLYF QGDFTVSPEN NETWQGAGVH ISEDSTVTWK
 451  VNGVANDRLS KIGKGTLHVQ AKGENQGSIS VGDGTVILDQ QADDKGKKQA
 501  FSEIGLVSGR GTVQLNADNQ FNPDKLYFGF RGGRLDLNGH SLSFHRIQNT
 551  DEGAMIVNHN QDKESTVTIT GNKDIATTGN NNSLDSKKEI AYNGWFGEKD
 601  TTKTNGRLNL VYQPAAEDRT LLLSGGTNLN GNITQTNGKL FFSGRPTPHA
 651  YNHLNDHWSQ KEGIPRGEIV WDNDWINRTF KAENFQIKGG QAVVSRNVAK
 701  VKGDWHLSNH AQAVFGVAPH QSHTICTRSD WTGLTNCVEK TITDDKVIAS
 751  LTKTDISGNV DLADHAHLNL TGLATLNGNL SANGDTRYTV SHNATQNGNL
 801  SLVGNAQATF NQATLNGNTS ASGNASFNLS DHAVQNGSLT LSGNAKANVS
 851  HSALNGNVSL ADKAVFHFES SRFTGQISGG KDTALHLKDS EWTLPSGTEL
 901  GNLNLDNATI TLNSAYRHDA AGAQTGSATD APRRRSRRSR RSLLSVTPPT
 951  SVESRFNTLT VNGKLNGQGT FRFMSELFGY RSDKLKLAES SEGTYTLAVN
1001  NTGNEPASLE QLTVVEGKDN KPLSENLNFT LQNEHVDAGA WRYQLIRKDG
1051  EFRLHNPVKE QELSDKLGKA EAKKQAEKDN AQSLDALIAA GRDAVEKTES
1101  VAEPARQAGG ENVGIMQAEE EKKRVQADKD TALAKQREAE TRPATTAFPR
1151  ARRARRDLPQ LQPQPQPQPQ RDLISRYANS GLSEFSATLN SVFAVQDELD
1201  RVFAEDRRNA VWTSGIRDTK HYRSQDFRAY RQQTDLRQIG MQKNLGSGRV
1251  GILFSHNRTE NTFDDGIGNS ARLAHGAVFG QYGIDRFYIG ISAGAGFSSG
1301  SLSDGIGGKI RRRVLHYGIQ ARYRAGFGGF GIEPHIGATR YFVQKADYRY
1351  ENVNIATPGL AFNRYRAGIK ADYSFKPAQH ISITPYLSLS YTDAASGKVR
1401  TRVNTAVLAQ DFGKTRSAEW GVNAEIKGFT LSLHAAAAKG PQLEAQHSAG
1451  IKLGYRW*
```

The leader peptide is underlined.

A polymorphic form of ORF1 is disclosed in WO99/55873.

Three expression strategies have been used for ORF1:

1) ORF1 using a His tag, following WO99/24578 (ORF1-His);

2) ORF1 with its own leader peptide but without any fusion partner ('ORF1L'); and

3) ORF1 with the leader peptide (MKKTAIAIAVALAGFATVAQA) from *E.coli* OmpA ('Orf1LOmpA'):

```
MKKTAIAIAVALAGFATVAQAASAGHTYFGINYQYYRDFAENKGKFAVGAKDIEVYNKKGELVGKSMTKAPMIDFSV
VSRNGVAALVGDQYIVSVAHNGGYNNVDFGAEGRNPDQHRFTYKIVKRNNYKAGTKGHPYGGDYHMPRLHKFVTDAE
PVEMTSYMDGRKYIDQNNYPDRVRIGAGRQYWRSDEDEPNNRESSYHIASAYSWLVGGNTFAQNGSGGGTVNLGSEK
IKHSPYGFLPTGGSFGDSGSPMFIYDAQKQKWLINGVLQTGNPYIGKSNGFQLVRKDWFYDEIFAGDTHSVFYEPRQ
NGKYSFNDDNNGTGKINAKHEHNSLPNRLKTRTVQLFNVSLSETAREPVYHAAGGVNSYRPRLNNGENISFIDEGKG
ELILTSNINQGAGGLYFQGDFTVSPENNETWQGAGVHISEDSTVTWKVNGVANDRLSKIGKGTLHVQAKGENQGSIS
VGDGTVILDQQADDKGKKQAFSEIGLVSGRGTVQLNADNQFNPDKLYFGFRGGRLDLNGHSLSFHRIQNTDEGAMIV
NHNQDKESTVTITGNKDIATTGNNNSLDSKKEIAYNGWFGEKDTTKTNGRLNLVYQPAAEDRTLLLSGGTNLNGNIT
QTNGKLFFSGRPTPHAYNHLNDHWSQKEGIPRGEIVWDNDWINRTFKAENFQIKGGQAVVSRNVAKVKGDWHLSNHA
QAVFGVAPHQSHTICTRSDWTGLTNCVEKTITDDKVIASLTKTDISGNVDLADHAHLNLTGLATLNGNLSANGDTRY
TVSHNATQNGNLSLVGNAQATFNQATLNGNTSASGNASFNLSDHAVQNGSLTLSGNAKANVSHSALNGNVSLADKAV
FHFESSRFTGQISGGKDTALHLKDSEWTLPSGTELGNLNLDNATITLNSAYRHDAAGAQTGSATDAPRRRSRRSRRS
LLSVTPPTSVESRFNTLTVNGKLNGQGTFRFMSELFGYRSDKLKLAESSEGTYTLAVNNTGNEPASLEQLTVVEGKD
NKPLSENLNFTLQNEHVDAGAWRYQLIRKDGEFRLHNPVKEQELSDKLGKAEAKKQAEKDNAQSLDALIAAGRDAVE
KTESVAEPARQAGGENVGIMQAEEEKKRVQADKDTALAKQREAETRPATTAFPRARRARRDLPQLQPQPQPQPQRDL
ISRYANSGLSEFSATLNSVFAVQDELDRVFAEDRRNAVWTSGIRDTKHYRSQDFRAYRQQTDLRQIGMQKNLGSGRV
GILFSHNRTENTFDDGIGNSARLAHGAVFGQYGIDRFYIGISAGAGFSSGSLSDGIGGKIRRRVLHYGIQARYRAGF
GGFGIEPHIGATRYFVQKADYRYENVNIATPGLAFNRYRAGIKADYSFKPAQHISITPYLSLSYTDAASGKVRTRVN
TAVLAQDFGKTRSAEWGVNAEIKGFTLSLHAAAAKGPQLEAQHSAGIKLGYRW*
```

To make this construct, the clone pET911LOmpA (see below) was digested with the *Nhe*I and *Xho*I restriction enzymes and the fragment corresponding to the vector carrying the OmpA leader sequence was purified (pETLOmpA). The ORF1 gene coding for the mature protein was amplified using the oligonucleotides ORF1-For and ORF1-Rev (including the *Nhe*I and *Xho*I restriction sites, respectively), digested with *Nhe*I and *Xho*I and ligated to the purified pETOmpA fragment (see Figure 1). An additional AS dipeptide was introduced by the *Nhe*I site.

All three forms of the protein were expressed. The His-tagged protein could be purified and was confirmed as surface exposed, and possibly secreted (see Figure 3). The protein was used to immunise mice, and the resulting sera gave excellent results in the bactericidal assay.

ORF1LOmpA was purified as total membranes, and was localised in both the inner and outer membranes. Unexpectedly, sera raised against ORF1LOmpA show even better ELISA and anti-bactericidal properties than those raised against the His-tagged protein.

ORF1L was purified as outer membranes, where it is localised.

Example 7 – protein 911 and its leader peptide

Protein 911 from *N.meningitidis* (serogroup B, strain MC58) has the following sequence:

```
  1  MKKNILEFWV GLFVLIGAAA VAFLAFRVAG GAAFGGSDKT YAVYADFGDI
 51  GGLKVNAPVK SAGVLVGRVG AIGLDPKSYQ ARVRLDLDGK YQFSSDVSAQ
101  ILTSGLLGEQ YIGLQQGGDT ENLAAGDTIS VTSSAMVLEN LIGKFMTSFA
151  EKNADGGNAE KAAE*
```

The leader peptide is underlined.

Three expression strategies have been used for 911:

1) 911 with its own leader peptide but without any fusion partner ('911L');
2) 911 with the leader peptide from *E.coli* OmpA ('911LOmpA').

   To make this construct, the entire sequence encoding the OmpA leader peptide was included in the 5'- primer as a tail (primer 911LOmpA Forward). A *Nhe*I restriction site was inserted between the sequence coding for the OmpA leader peptide and the 911 gene encoding the predicted mature protein (insertion of one amino acid, a serine), to allow the use of this construct to clone different genes downstream the OmpA leader peptide sequence.
3) 911 with the leader peptide (MKYLLPTAAAGLLLAAQPAMA) from *Erwinia carotovora* PelB ('911LpelB').

To make this construct, the 5'-end PCR primer was designed downstream from the leader sequence and included the *Nco*I restriction site in order to have the 911 fused directly to the PelB leader sequence; the 3'- end primer included the STOP codon. The expression vector used was pET22b+ (Novagen), which carries the coding sequence for the PelB leader peptide. The *Nco*I site introduces an additional methionine after the PelB sequence.

All three forms of the protein were expressed. ELISA titres were highest using 911L, with 919LOmpA also giving good results.

Example 8 – ORF46

The complete ORF46 protein from *N.meningitidis* (serogroup B, strain 2996) has the following sequence:

```
  1  LGISRKISLI LSILAVCLPM HAHASDLAND SFIRQVLDRQ HFEPDGKYHL
 51  FGSRGELAER SGHIGLGKIQ SHQLGNLMIQ QAAIKGNIGY IVRFSDHGHE
101  VHSPFDNHAS HSDSDEAGSP VDGFSLYRIH WDGYEHHPAD GYDGPQGGGY
151  PAPKGARDIY SYDIKGVAQN IRLNLTDNRS TGQRLADRFH NAGSMLTQGV
201  GDGFKRATRY SPELDRSGNA AEAFNGTADI VKNIIGAAGE IVGAGDAVQG
251  ISEGSNIAVM HGLGLLSTEN KMARINDLAD MAQLKDYAAA AIRDWAVQNP
301  NAAQGIEAVS NIFMAAIPIK GIGAVRGKYG LGGITAHPIK RSQMGAIALP
351  KGKSAVSDNF ADAAYAKYPS PYHSRNIRSN LEQRYGKENI TSSTVPPSNG
401  KNVKLADQRH PKTGVPFDGK GFPNFEKHVK YDTKLDIQEL SGGGIPKAKP
451  VSDAKPRWEV DRKLNKLTTR EQVEKNVQEI RNGNKNSNFS QHAQLEREIN
501  KLKSADEINF ADGMGKFTDS MNDKAFSRLV KSVKENGFTN PVVEYVEING
551  KAYIVRGNNR VFAAEYLGRI HELKFKKVDF PVPNTSWKNP TDVLNESGNV
601  KRPRYRSK*
```

The leader peptide is underlined.

The sequences of ORF46 from other strains can be found in WO00/66741.

Three expression strategies have been used for ORF46:

1) ORF46 with its own leader peptide but without any fusion partner ('ORF46-2L');
2) ORF46 without its leader peptide and without any fusion partner ('ORF46-2'), with the leader peptide omitted by designing the 5'-end amplification primer downstream from the predicted leader sequence:

```
  1  SDLANDSFIR QVLDRQHFEP DGKYHLFGSR GELAERSGHI GLGKIQSHQL
 51  GNLMIQQAAI KGNIGYIVRF SDHGHEVHSP FDNHASHSDS DEAGSPVDGF
101  SLYRIHWDGY EHHPADGYDG PQGGGYPAPK GARDIYSYDI KGVAQNIRLN
151  LTDNRSTGQR LADRFHNAGS MLTQGVGDGF KRATRYSPEL DRSGNAAEAF
201  NGTADIVKNI IGAAGEIVGA GDAVQGISEG SNIAVMHGLG LLSTENKMAR
251  INDLADMAQL KDYAAAAIRD WAVQNPNAAQ GIEAVSNIFM AAIPIKGIGA
301  VRGKYGLGGI TAHPIKRSQM GAIALPKGKS AVSDNFADAA YAKYPSPYHS
351  RNIRSNLEQR YGKENITSST VPPSNGKNVK LADQRHPKTG VPFDGKGFPN
401  FEKHVKYDTK LDIQELSGGG IPKAKPVSDA KPRWEVDRKL NKLTTREQVE
451  KNVQEIRNGN KNSNFSQHAQ LEREINKLKS ADEINFADGM GKFTDSMNDK
501  AFSRLVKSVK ENGFTNPVVE YVEINGKAYI VRGNNRVFAA EYLGRIHELK
551  FKKVDFPVPN TSWKNPTDVL NESGNVKRPR YRSK*
```

3) ORF46 as a truncated protein, consisting of the first 433 amino acids ('ORF46.1L'), constructed by designing PCR primers to amplify a partial sequence corresponding to aa 1-433.

A STOP codon was included in the 3'-end primer sequences.

ORF46-2L is expressed at a very low level to *E.coli*. Removal of its leader peptide (ORF46-2) does not solve this problem. The truncated ORF46.1L form (first 433 amino acids, which are well conserved between serogroups and species), however, is well-expressed and gives excellent results in ELISA test and in the bactericidal assay.

ORF46.1 has also been used as the basis of hybrid proteins. It has been fused with 287, 919, and ORF1. The hybrid proteins were generally insoluble, but gave some good ELISA and bactericidal results (against the homologous 2996 strain):

Protein	ELISA	Bactericidal Ab
Orf1-Orf46.1-His	850	256
919-Orf46.1-His	12900	512
919-287-Orf46-His	n.d.	n.d.
Orf46.1-287His	150	8192
Orf46.1-919His	2800	2048
Orf46.1-287-919His	3200	16384

For comparison, 'triple' hybrids of ORF46.1, 287 (either as a GST fusion, or in ΔG287 form) and 919 were constructed and tested against various strains (including the homologous 2996 strain) *versus* a simple mixture of the three antigens. FCA was used as adjuvant:

	2996	**BZ232**	**MC58**	**NGH38**	**F6124**	**BZ133**
Mixture	8192	256	512	1024	>2048	>2048
ORF46.1-287-919his	16384	256	4096	8192	8192	8192
ΔG287-919-ORF46.1his	8192	64	4096	8192	8192	16384
ΔG287-ORF46.1-919his	4096	128	256	8192	512	1024

Again, the hybrids show equivalent or superior immunological activity.

Hybrids of two proteins (strain 2996) were compared to the individual proteins against various heterologous strains:

	1000	MC58	F6124 (MenA)
ORF46.1-His	<4	4096	<4
ORF1-His	8	256	128
ORF1—ORF46.1-His	1024	512	1024

Again, the hybrid shows equivalent or superior immunological activity.

Example 9 – protein 961

The complete 961 protein from *N.meningitidis* (serogroup B, strain MC58) has the following sequence:

```
  1  MSMKHFPAKV LTTAILATFC SGALAATSDD DVKKAATVAI VAAYNNGQEI
 51  NGFKAGETIY DIGEDGTITQ KDATAADVEA DDFKGLGLKK VVTNLTKTVN
101  ENKQNVDAKV KAAESEIEKL TTKLADTDAA LADTDAALDE TTNALNKLGE
151  NITTFAEETK TNIVKIDEKL EAVADTVDKH AEAFNDIADS LDETNTKADE
201  AVKTANEAKQ TAEETKQNVD AKVKAAETAA GKAEAAAGTA NTAADKAEAV
251  AAKVTDIKAD IATNKADIAK NSARIDSLDK NVANLRKETR QGLAEQAALS
301  GLFQPYNVGR FNVTAAVGGY KSESAVAIGT GFRFTENFAA KAGVAVGTSS
351  GSSAAYHVGV NYEW*
```

The leader peptide is underlined.

Three approaches to 961 expression were used:

1) 961 using a GST fusion, following WO99/57280 ('GST961');
2) 961 with its own leader peptide but without any fusion partner ('961L'); and
3) 961 without its leader peptide and without any fusion partner ('$961^{untagged}$'), with the leader peptide omitted by designing the 5'-end PCR primer downstream from the predicted leader sequence.

All three forms of the protein were expressed. The GST-fusion protein could be purified and antibodies against it confirmed that 961 is surface exposed (Figure 4). The protein was used to immunise mice, and the resulting sera gave excellent results in the bactericidal assay. 961L could also be purified and gave very high ELISA titres.

Protein 961 appears to be phase variable. Furthermore, it is not found in all strains of *N.meningitidis*.

Example 10 – protein 287

Protein 287 from *N.meningitidis* (serogroup B, strain 2996) has the following sequence:

```
  1  MFERSVIAMA CIFALSACGG GGGGSPDVKS ADTLSKPAAP VVAEKETEVK
 51  EDAPQAGSQG QGAPSTQGSQ DMAAVSAENT GNGGAATTDK PKNEDEGPQN
101  DMPQNSAESA NQTGNNQPAD SSDSAPASNP APANGGSNFG RVDLANGVLI
151  DGPSQNITLT HCKGDSCNGD NLLDEEAPSK SEFENLNESE RIEKYKKDGK
```

```
201  SDKFTNLVAT AVQANGTNKY VIIYKDKSAS SSSARFRRSA RSRRSLPAEM
251  PLIPVNQADT LIVDGEAVSL TGHSGNIFAP EGNYRYLTYG AEKLPGGSYA
301  LRVQGEPAKG EMLAGTAVYN GEVLHFHTEN GRPYPTRGRF AAKVDFGSKS
351  VDGIIDSGDD LHMGTQKFKA AIDGNGFKGT WTENGGGDVS GRFYGPAGEE
401  VAGKYSYRPT DAEKGGFGVF AGKKEQD*
```

The leader peptide is shown underlined.

The sequences of 287 from other strains can be found in Figures 5 and 15 of WO00/66741.

Example 9 of WO99/57280 discloses the expression of 287 as a GST-fusion in *E.coli*.

A number of further approaches to expressing 287 in *E.coli* have been used, including:

1) 287 as a His-tagged fusion ('287-His');
2) 287 with its own leader peptide but without any fusion partner ('287L');
3) 287 with the ORF4 leader peptide and without any fusion partner ('287LOrf4'); and
4) 287 without its leader peptide and without any fusion partner ('$287^{untagged}$'):

```
  1  CGGGGGGSPD VKSADTLSKP AAPVVAEKET EVKEDAPQAG SQGQGAPSTQ
 51  GSQDMAAVSA ENTGNGGAAT TDKPKNEDEG PQNDMPQNSA ESANQTGNNQ
101  PADSSDSAPA SNPAPANGGS NFGRVDLANG VLIDGPSQNI TLTHCKGDSC
151  NGDNLLDEEA PSKSEFENLN ESERIEKYKK DGKSDKFTNL VATAVQANGT
201  NKYVIIYKDK SASSSSARFR RSARSRRSLP AEMPLIPVNQ ADTLIVDGEA
251  VSLTGHSGNI FAPEGNYRYL TYGAEKLPGG SYALRVQGEP AKGEMLAGTA
301  VYNGEVLHFH TENGRPYPTR GRFAAKVDFG SKSVDGIIDS GDDLHMGTQK
351  FKAAIDGNGF KGTWTENGGG DVSGRFYGPA GEEVAGKYSY RPTDAEKGGF
401  GVFAGKKEQD *
```

All these proteins could be expressed and purified.

'287L' and '287LOrf4' were confirmed as lipoproteins.

As shown in Figure 2, '287LOrf4' was constructed by digesting 919LOrf4 with *Nhe*I and *Xho*I. The entire ORF4 leader peptide was restored by the addition of a DNA sequence coding for the missing amino acids, as a tail, in the 5'-end primer (287LOrf4 for), fused to 287 coding sequence. The 287 gene coding for the mature protein was amplified using the oligonucleotides 287LOrf4 For and Rev (including the *Nhe*I and *Xho*I sites, respectively), digested with *Nhe*I and *Xho*I and ligated to the purified pETOrf4 fragment.

Example 11 – further non-fusion proteins with/without native leader peptides

A similar approach was adopted for *E.coli* expression of further proteins from WO99/24578, WO99/36544 and WO99/57280.

The following were expressed without a fusion partner: 008, 105, 117-1, 121-1, 122-1, 128-1, 148, 216, 243, 308, 593, 652, 726, 982, and Orf143-1. Protein 117-1 was confirmed as surface-exposed by FACS and gave high ELISA titres.

The following were expressed with the native leader peptide but without a fusion partner: 111, 149, 206, 225-1, 235, 247-1, 274, 283, 286, 292, 401, 406, 502-1, 503, 519-1, 525-1, 552, 556, 557, 570, 576-1, 580, 583, 664, 759, 907, 913, 920-1, 926, 936-1, 953, 961, 983, 989, Orf4, Orf7-1, Orf9-1, Orf23, Orf25, Orf37, Orf38, Orf40, Orf40.1, Orf40.2, Orf72-1, Orf76-1, Orf85-2, Orf91, Orf97-1, Orf119, Orf143.1. These proteins are given the suffix 'L'.

His-tagged protein 760 was expressed with and without its leader peptide. The deletion of the signal peptide greatly increased expression levels. The protein could be purified most easily using 2M urea for solubilisation.

His-tagged protein 264 was well-expressed using its own signal peptide, and the 30kDa protein gave positive Western blot results.

All proteins were successfully expressed.

The localisation of 593, 121-1, 128-1, 593, 726, and 982 in the cytoplasm was confirmed.

The localisation of 920-1L, 953L, ORF9-1L, ORF85-2L, ORF97-1L, 570L, 580L and 664L in the periplasm was confirmed.

The localisation of ORF40L in the outer membrane, and 008 and 519-1L in the inner membrane was confirmed. ORF25L, ORF4L, 406L, 576-1L were all confirmed as being localised in the membrane.

Protein 206 was found not to be a lipoprotein.

ORF25 and ORF40 expressed with their native leader peptides but without fusion partners, and protein 593 expressed without its native leader peptide and without a fusion partner, raised good anti-bactericidal sera. Surprisingly, the forms of ORF25 and ORF40 expressed without fusion partners and using their own leader peptides (*i.e.* 'ORF25L' and 'ORF40L') give better results in the bactericidal assay than the fusion proteins.

Proteins 920L and 953L were subjected to N-terminal sequencing, giving HRVWVETAH and ATYKVDEYHANARFAF, respectively. This sequencing confirms that the predicted leader peptides were cleaved and, when combined with the periplasmic location, confirms that the

proteins are correctly processed and localised by *E.coli* when expressed from their native leader peptides.

The N-terminal sequence of protein 519.1L localised in the inner membrane was MEFFIILLA, indicating that the leader sequence is not cleaved. It may therefore function as both an uncleaved leader sequence and a transmembrane anchor in a manner similar to the leader peptide of PBP1 from *N.gonorrhoeae* [Ropp & Nicholas (1997) *J. Bact.* 179:2783-2787.]. Indeed the N-terminal region exhibits strong hydrophobic character and is predicted by the Tmpred. program to be transmembrane.

Example 12 – lipoproteins

The incorporation of palmitate in recombinant lipoproteins was demonstrated by the method of Kraft *et. al.* [*J. Bact.* (1998) 180:3441-3447.]. Single colonies harbouring the plasmid of interest were grown overnight at 37°C in 20 ml of LB/Amp (100µg/ml) liquid culture. The culture was diluted to an OD_{550} of 0.1 in 5.0 ml of fresh medium LB/Amp medium containing 5 µC/ml [^{3}H] palmitate (Amersham). When the OD_{550} of the culture reached 0.4-0.8, recombinant lipoprotein was induced for 1 hour with IPTG (final concentration 1.0 mM). Bacteria were harvested by centrifugation in a bench top centrifuge at 2700*g* for 15 min and washed twice with 1.0 ml cold PBS. Cells were resuspended in 120µl of 20 mM Tris-HCl (pH 8.0), 1 mM EDTA, 1.0% w/v SDS and lysed by boiling for 10 min. After centrifugation at 13000*g* for 10 min the supernatant was collected and proteins precipitated by the addition of 1.2 ml cold acetone and left for 1 hour at -20 °C. Protein was pelleted by centrifugation at 13000*g* for 10 min and resuspended in 20-50µl (calculated to standardise loading with respect to the final O.D of the culture) of 1.0% w/v SDS. An aliquot of 15 µl was boiled with 5µl of SDS-PAGE sample buffer and analysed by SDS-PAGE. After electrophoresis gels were fixed for 1 hour in 10% v/v acetic acid and soaked for 30 minutes in Amplify solution (Amersham). The gel was vacuum-dried under heat and exposed to Hyperfilm (Kodak) overnight -80 °C.

Incorporation of the [^{3}H] palmitate label, confirming lipidation, was found for the following proteins: Orf4L, Orf25L, 287L, 287LOrf4, 406.L, 576L, 926L, 919L and 919LOrf4.

Example 13 – domains in 287

Based on homology of different regions of 287 to proteins that belong to different functional classes, it was split into three 'domains', as shown in Figure 5. The second domain shows

homology to IgA proteases, and the third domain shows homology to transferrin-binding proteins.

Each of the three 'domains' shows a different degree of sequence conservation between *N.meningitidis* strains – domain C is 98% identical, domain A is 83% identical, whilst domain B is only 71% identical. Note that protein 287 in strain MC58 is 61 amino acids longer than that of strain 2996. An alignment of the two sequences is shown in Figure 7, and alignments for various strains are disclosed in WO00/66741 (see Figures 5 and 15 therein).

The three domains were expressed individually as C-terminal His-tagged proteins. This was done for the MC58 and 2996 strains, using the following constructs:

287a-MC58 (aa 1-202), 287b-MC58 (aa 203-288), 287c-MC58 (aa 311-488).

287a-2996 (aa 1-139), 287b-2996 (aa 140-225), 287c-2996 (aa 250-427).

To make these constructs, the stop codon sequence was omitted in the 3'-end primer sequence. The 5' primers included the *Nhe*I restriction site, and the 3' primers included a *Xho*I as a tail, in order to direct the cloning of each amplified fragment into the expression vector pET21b+ using *Nde*I-*Xho*I, *Nhe*I-*Xho*I or *Nde*I-*Hind*III restriction sites.

All six constructs could be expressed, but 287b-MC8 required denaturation and refolding for solubilisation.

Deletion of domain A is described below ('Δ4 287-His').

Immunological data (serum bactericidal assay) were also obtained using the various domains from strain 2996, against the homologous and heterologous MenB strains, as well as MenA (F6124 strain) and MenC (BZ133 strain):

	2996	**BZ232**	**MC58**	**NGH38**	**394/98**	**MenA**	**MenC**
287-His	32000	16	4096	4096	512	8000	16000
287(B)-His	256	-	-	-	-	16	-
287(C)-His	256	-	32	512	32	2048	>2048
287(B-C)-His	64000	128	4096	64000	1024	64000	32000

Using the domains of strain MC58, the following results were obtained: .

	MC58	2996	BZ232	NGH38	394/98	MenA	MenC
287-His	4096	32000	16	4096	512	8000	16000
287(B)-His	128	128	-	-	-	-	128
287(C)-His	-	16	-	1024	-	512	-
287(B-C)-His	16000	64000	128	64000	512	64000	>8000

Example 14 – deletions in 287

As well as expressing individual domains, 287 was also expressed (as a C-terminal His-tagged protein) by making progressive deletions within the first domain. These

Four deletion mutants of protein 287 from strain 2996 were used (Figure 6):

1) '287-His', consisting of amino acids 18-427 (*i.e.* leader peptide deleted);
2) 'Δ1 287-His', consisting of amino acids 26-427;
3) 'Δ2 287-His', consisting of amino acids 70-427;
4) 'Δ3 287-His', consisting of amino acids 107-427; and
5) 'Δ4 287-His', consisting of amino acids 140-427 (=287-bc).

The 'Δ4' protein was also made for strain MC58 ('Δ4 287MC58-His'; aa 203-488).

The constructs were made in the same way as 287a/b/c, as described above.

All six constructs could be expressed and protein could be purified. Expression of 287-His was, however, quite poor.

Expression was also high when the C-terminal His-tags were omitted.

Immunological data (serum bactericidal assay) were also obtained using the deletion mutants, against the homologous (2996) and heterologous MenB strains, as well as MenA (F6124 strain) and MenC (BZ133 strain):

	2996	BZ232	MC58	NGH38	394/98	MenA	MenC
287-his	32000	16	4096	4096	512	8000	16000
Δ1 287-His	16000	128	4096	4096	1024	8000	16000
Δ2 287-His	16000	128	4096	>2048	512	16000	>8000
Δ3 287-His	16000	128	4096	>2048	512	16000	>8000
Δ4 287-His	64000	128	4096	64000	1024	64000	32000

The same high activity for the Δ4 deletion was seen using the sequence from strain MC58.

As well as showing superior expression characteristics, therefore, the mutants are immunologically equivalent or superior.

Example 15 – poly-glycine deletions

The 'Δ1 287-His' construct of the previous example differs from 287-His and from '287^{untagged}' only by a short N-terminal deletion (GGGGGGS). Using an expression vector which replaces the deleted serine with a codon present in the *Nhe* cloning site, however, this amounts to a deletion only of $(Gly)_6$. Thus, the deletion of this $(Gly)_6$ sequence has been shown to have a dramatic effect on protein expression.

The protein lacking the N-terminal amino acids up to GGGGGG is called 'ΔG 287'. In strain MC58, its sequence (leader peptide underlined) is:

```
                                        ➦ ΔG287
  1  MFKRSVIAMA CIFALSACGG GGGGSPDVKS ADTLSKPAAP VVSEKETEAK
 51  EDAPQAGSQG QGAPSAQGSQ DMAAVSEENT GNGGAVTADN PKNEDEVAQN
101  DMPQNAAGTD SSTPNHTPDP NMLAGNMENQ ATDAGESSQP ANQPDMANAA
151  DGMQGDDPSA GGQNAGNTAA QGANQAGNNQ AAGSSDPIPA SNPAPANGGS
201  NFGRVDLANG VLIDGPSQNI TLTHCKGDSC SGNNFLDEEV QLKSEFEKLS
251  DADKISNYKK DGKNDKFVGL VADSVQMKGI NQYIIFYKPK PTSFARFRRS
301  ARSRRSLPAE MPLIPVNQAD TLIVDGEAVS LTGHSGNIFA PEGNYRYLTY
351  GAEKLPGGSY ALRVQGEPAK GEMLAGAAVY NGEVLHFHTE NGRPYPTRGR
401  FAAKVDFGSK SVDGIIDSGD DLHMGTQKFK AAIDGNGFKG TWTENGSGDV
451  SGKFYGPAGE EVAGKYSYRP TDAEKGGFGV FAGKKEQD*
```

ΔG287, with or without His-tag ('ΔG287-His' and 'ΔG287K', respectively), are expressed at very good levels in comparison with the '287-His' or '287 ^{untagged}'.

On the basis of gene variability data, variants of ΔG287-His were expressed in *E.coli* from a number of MenB strains, in particular from strains 2996, MC58, 1000, and BZ232. The results were also good.

It was hypothesised that poly-Gly deletion might be a general strategy to improve expression. Other MenB lipoproteins containing similar $(Gly)_n$ motifs (near the N-terminus, downstream of a cysteine) were therefore identified, namely Tbp2 (NMB0460), 741 (NMB 1870) and 983 (NMB1969):

```
TBP2                                    ➦ ΔGTbp2
  1  MNNPLVNQAA MVLPVFLLSA CLGGGGSFDL DSVDTRAPRP APKYQDVPSE
 51  KPQAQKDQGG YGFAMRLKRR NWYPQAKEDE VKLDESDWEA TGLPDEPKEL
101  PKRQKSVIEK VETDSDNNIY SSPYLKPSNH QNGNTGNGIN QPKNQAKDYE
151  NFKYVYSGWF YKHAKREFNL KVEPKSAKNG DDGYIFYHGK EPSRQLPASG
201  KITYKGVWHF ATDTKKGQKF REIIQPSKSQ GDRYSGFSGD DGEEYSNKNK
251  STLTDGQEGY GFTSNLEVDF HNKKLTGKLI RNNANTDNNQ ATTTQYYSLE
301  AQVTGNRFNG KATATDKPQQ NSETKEHPFV SDSSSLSGGF FGPQGEELGF
351  RFLSDDQKVA VVGSAKTKDK PANGNTAAAS GGTDAAASNG AAGTSSENGK
401  LTTVLDAVEL KLGDKEVQKL DNFSNAAQLV VDGIMIPLLP EASESGNNQA
451  NQGTNGGTAF TRKFDHTPES DKKDAQAGTQ TNGAQTASNT AGDTNGKTKT
```

```
 501  YEVEVCCSNL NYLKYGMLTR KNSKSAMQAG ESSSQADAKT EQVEQSMFLQ
 551  GERTDEKEIP SEQNIVYRGS WYGYIANDKS TSWSGNASNA TSGNRAEFTV
 601  NFADKKITGT LTADNRQEAT FTIDGNIKDN GFEGTAKTAE SGFDLDQSNT
 651  TRTPKAYITD AKVQGGFYGP KAEELGGWFA YPGDKQTKNA TNASGNSSAT
 701  VVFGAKRQQP VR*

741                                  ➡ ΔG741
   1  VNRTAFCCLS LTTALILTAC SSGGGGVAAD IGAGLADALT APLDHKDKGL
  51  QSLTLDQSVR KNEKLKLAAQ GAEKTYGNGD SLNTGKLKND KVSRFDFIRQ
 101  IEVDGQLITL ESGEFQVYKQ SHSALTAFQT EQIQDSEHSG KMVAKRQFRI
 151  GDIAGEHTSF DKLPEGGRAT YRGTAFGSDD AGGKLTYTID FAAKQGNGKI
 201  EHLKSPELNV DLAAADIKPD GKRHAVISGS VLYNQAEKGS YSLGIFGGKA
 251  QEVAGSAEVK TVNGIRHIGL AAKQ*

983                                                  ➡ ΔG983
   1  MRTTPTFPTK TFKPTAMALA VATTLSACLG GGGGGTSAPD FNAGGTGIGS
  51  NSRATTAKSA AVSYAGIKNE MCKDRSMLCA GRDDVAVTDR DAKINAPPPN
 101  LHTGDFPNPN DAYKNLINLK PAIEAGYTGR GVEVGIVDTG ESVGSISFPE
 151  LYGRKEHGYN ENYKNYTAYM RKEAPEDGGG KDIEASFDDE AVIETEAKPT
 201  DIRHVKEIGH IDLVSHIIGG RSVDGRPAGG IAPDATLHIM NTNDETKNEM
 251  MVAAIRNAWV KLGERGVRIV NNSFGTTSRA GTADLFQIAN SEEQYRQALL
 301  DYSGGDKTDE GIRLMQQSDY GNLSYHIRNK NMLFIFSTGN DAQAQPNTYA
 351  LLPFYEKDAQ KGIITVAGVD RSGEKFKREM YGEPGTEPLE YGSNHCGITA
 401  MWCLSAPYEA SVRFTRTNPI QIAGTSFSAP IVTGTAALLL QKYPWMSNDN
 451  LRTTLLTTAQ DIGAVGVDSK FGWGLLDAGK AMNGPASFPF GDFTADTKGT
 501  SDIAYSFRND ISGTGGLIKK GGSQLQLHGN NTYTGKTIIE GGSLVLYGNN
 551  KSDMRVETKG ALIYNGAASG GSLNSDGIVY LADTDQSGAN ETVHIKGSLQ
 601  LDGKGTLYTR LGKLLKVDGT AIIGGKLYMS ARGKGAGYLN STGRRVPFLS
 651  AAKIGQDYSF FTNIETDGGL LASLDSVEKT AGSEGDTLSY YVRRGNAART
 701  ASAAAHSAPA GLKHAVEQGG SNLENLMVEL DASESSATPE TVETAAADRT
 751  DMPGIRPYGA TFRAAAAVQH ANAADGVRIF NSLAATVYAD STAAHADMQG
 801  RRLKAVSDGL DHNGTGLRVI AQTQQDGGTW EQGGVEGKMR GSTQTVGIAA
 851  KTGENTTAAA TLGMGRSTWS ENSANAKTDS ISLFAGIRHD AGDIGYLKGL
 901  FSYGRYKNSI SRSTGADEHA EGSVNGTLMQ LGALGGVNVP FAATGDLTVE
 951  GGLRYDLLKQ DAFAEKGSAL GWSGNSLTEG TLVGLAGLKL SQPLSDKAVL
1001  FATAGVERDL NGRDYTVTGG FTGATAATGK TGARNMPHTR LVAGLGADVE
1051  FGNGWNGLAR YSYAGSKQYG NHSGRVGVGY RF*
```

Tbp2 and 741 genes were from strain MC58; 983 and 287 genes were from strain 2996. These were cloned in pET vector and expressed in *E.coli* without the sequence coding for their leader peptides or as "ΔG forms", both fused to a C-terminal His-tag. In each case, the same effect was seen – expression was good in the clones carrying the deletion of the poly-glycine stretch, and poor or absent if the glycines were present in the expressed protein:

ORF	Express.	Purification	Bact. Activity
287-His(2996)	+/-	+	+
'287^{untagged}'(2996)	+/-	nd	nd
ΔG287-His(2996)	+	+	+
ΔG287K(2996)	+	+	+
ΔG287-His(MC58)	+	+	+
ΔG287-His(1000)	+	+	+
ΔG287-His(BZ232)	+	+	+
Tbp2-His(MC58)	+/-	nd	nd
ΔGTbp2-His(MC58)	+	+	
741-His(MC58)	+/-	nd	nd
ΔG741-His(MC58)	+	+	
983-His (2996)			
ΔG983-His (2996)	+	+	

SDS-PAGE of the proteins is shown in Figure 13.

ΔG287 and hybrids

ΔG287 proteins were made and purified for strains MC58, 1000 and BZ232. Each of these gave high ELISA titres and also serum bactericidal titres of >8192. ΔG287K, expressed from pET-24b, gave excellent titres in ELISA and the serum bactericidal assay. ΔG287-ORF46.1K may also be expressed in pET-24b.

ΔG287 was also fused directly in-frame upstream of 919, 953, 961 (sequences shown below) and ORF46.1:

ΔG287-919

```
   1  ATGGCTAGCC CCGATGTTAA ATCGGCGGAC ACGCTGTCAA AACCGGCCGC
  51  TCCTGTTGTT GCTGAAAAAG AGACAGAGGT AAAAGAAGAT GCGCCACAGG
 101  CAGGTTCTCA AGGACAGGGC GCGCCATCCA CACAAGGCAG CCAAGATATG
 151  GCGGCAGTTT CGGCAGAAAA TACAGGCAAT GGCGGTGCGG CAACAACGGA
 201  CAAACCCAAA AATGAAGACG AGGGACCGCA AAATGATATG CCGCAAAATT
 251  CCGCCGAATC CGCAAATCAA ACAGGGAACA ACCAACCCGC CGATTCTTCA
 301  GATTCCGCCC CCGCGTCAAA CCCTGCACCT GCGAATGGCG GTAGCAATTT
 351  TGGAAGGGTT GATTTGGCTA ATGGCGTTTT GATTGATGGG CCGTCGCAAA
 401  ATATAACGTT GACCCACTGT AAAGGCGATT CTTGTAATGG TGATAATTTA
 451  TTGGATGAAG AAGCACCGTC AAAATCAGAA TTTGAAAATT TAAATGAGTC
 501  TGAACGAATT GAGAAATATA AGAAAGATGG GAAAAGCGAT AAATTTACTA
 551  ATTTGGTTGC GACAGCAGTT CAAGCTAATG GAACTAACAA ATATGTCATC
 601  ATTTATAAAG ACAAGTCCGC TTCATCTTCA TCTGCGCGAT TCAGGCGTTC
 651  TGCACGGTCG AGGAGGTCGC TTCCTGCCGA GATGCCGCTA ATCCCCGTCA
 701  ATCAGGCGGA TACGCTGATT GTCGATGGGG AAGCGGTCAG CCTGACGGGG
 751  CATTCCGGCA ATATCTTCGC GCCCGAAGGG AATTACCGGT ATCTGACTTA
 801  CGGGGCGGAA AAATTGCCCG GCGGATCGTA TGCCCTCCGT GTGCAAGGCG
 851  AACCGGCAAA AGGCGAAATG CTTGCTGGCA CGGCCGTGTA CAACGGCGAA
 901  GTGCTGCATT TTCATACGGA AAACGGCCGT CCGTACCCGA CTAGAGGCAG
 951  GTTTGCCGCA AAAGTCGATT TCGGCAGCAA ATCTGTGGAC GGCATTATCG
1001  ACAGCGGCGA TGATTTGCAT ATGGGTACGC AAAAATTCAA AGCCGCCATC
```

```
1051  GATGGAAACG GCTTTAAGGG GACTTGGACG GAAAATGGCG GCGGGGATGT
1101  TTCCGGAAGG TTTTACGGCC CGGCCGGCGA GGAAGTGGCG GGAAAATACA
1151  GCTATCGCCC GACAGATGCG GAAAAGGGCG GATTCGGCGT GTTTGCCGGC
1201  AAAAAAGAGC AGGATGGATC CGGAGGAGGA GGATGCCAAA GCAAGAGCAT
1251  CCAAACCTTT CCGCAACCCG ACACATCCGT CATCAACGGC CCGGACCGGC
1301  CGGTCGGCAT CCCCGACCCC GCCGGAACGA CGGTCGGCGG CGGCGGGGCC
1351  GTCTATACCG TTGTACCGCA CCTGTCCCTG CCCCACTGGG CGGCGCAGGA
1401  TTTCGCCAAA AGCCTGCAAT CCTTCCGCCT CGGCTGCGCC AATTTGAAAA
1451  ACCGCCAAGG CTGGCAGGAT GTGTGCGCCC AAGCCTTTCA AACCCCCGTC
1501  CATTCCTTTC AGGCAAAACA GTTTTTTGAA CGCTATTTCA CGCCGTGGCA
1551  GGTTGCAGGC AACGGAAGCC TTGCCGGTAC GGTTACCGGC TATTACGAGC
1601  CGGTGCTGAA GGGCGACGAC AGGCGGACGG CACAAGCCCG CTTCCCGATT
1651  TACGGTATTC CCGACGATTT TATCTCCGTC CCCCTGCCTG CCGGTTTGCG
1701  GAGCGGAAAA GCCCTTGTCC GCATCAGGCA GACGGGAAAA AACAGCGGCA
1751  CAATCGACAA TACCGGCGGC ACACATACCG CCGACCTCTC CCGATTCCCC
1801  ATCACCGCGC GCACAACGGC AATCAAAGGC AGGTTTGAAG GAAGCCGCTT
1851  CCTCCCCTAC CACACGCGCA ACCAAATCAA CGGCGGCGCG CTTGACGGCA
1901  AAGCCCCGAT ACTCGGTTAC GCCGAAGACC CCGTCGAACT TTTTTTTATG
1951  CACATCCAAG GCTCGGGCCG TCTGAAAACC CCGTCCGGCA AATACATCCG
2001  CATCGGCTAT GCCGACAAAA ACGAACATCC CTACGTTTCC ATCGGACGCT
2051  ATATGGCGGA CAAAGGCTAC CTCAAGCTCG GGCAGACCTC GATGCAGGGC
2101  ATCAAAGCCT ATATGCGGCA AAATCCGCAA CGCCTCGCCG AAGTTTTGGG
2151  TCAAAACCCC AGCTATATCT TTTTCCGCGA GCTTGCCGGA AGCAGCAATG
2201  ACGGTCCCGT CGGCGCACTG GGCACGCCGT TGATGGGGGA ATATGCCGGC
2251  GCAGTCGACC GGCACTACAT TACCTTGGGC GCGCCCTTAT TTGTCGCCAC
2301  CGCCCATCCG GTTACCCGCA AAGCCCTCAA CCGCCTGATT ATGGCGCAGG
2351  ATACCGGCAG CGCGATTAAA GGCGCGGTGC GCGTGGATTA TTTTTGGGGA
2401  TACGGCGACG AAGCCGGCGA ACTTGCCGGC AAACAGAAAA CCACGGGTTA
2451  CGTCTGGCAG CTCCTACCCA ACGGTATGAA GCCCGAATAC CGCCCGTAAC
2501  TCGAG

   1  MASPDVKSAD TLSKPAAPVV AEKETEVKED APQAGSQGQG APSTQGSQDM
  51  AAVSAENTGN GGAATTDKPK NEDEGPQNDM PQNSAESANQ TGNNQPADSS
 101  DSAPASNPAP ANGGSNFGRV DLANGVLIDG PSQNITLTHC KGDSCNGDNL
 151  LDEEAPSKSE FENLNESERI EKYKKDGKSD KFTNLVATAV QANGTNKYVI
 201  IYKDKSASSS SARFRRSARS RRSLPAEMPL IPVNQADTLI VDGEAVSLTG
 251  HSGNIFAPEG NYRYLTYGAE KLPGGSYALR VQGEPAKGEM LAGTAVYNGE
 301  VLHFHTENGR PYPTRGRFAA KVDFGSKSVD GIIDSGDDLH MGTQKFKAAI
 351  DGNGFKGTWT ENGGGDVSGR FYGPAGEEVA GKYSYRPTDA EKGGFGVFAG
 401  KKEQDGSGGG GCQSKSIQTF PQPDTSVING PDRPVGIPDP AGTTVGGGGA
 451  VYTVVPHLSL PHWAAQDFAK SLQSFRLGCA NLKNRQGWQD VCAQAFQTPV
 501  HSFQAKQFFE RYFTPWQVAG NGSLAGTVTG YYEPVLKGDD RRTAQARFPI
 551  YGIPDDFISV PLPAGLRSGK ALVRIRQTGK NSGTIDNTGG THTADLSRFP
 601  ITARTTAIKG RFEGSRFLPY HTRNQINGGA LDGKAPILGY AEDPVELFFM
 651  HIQGSGRLKT PSGKYIRIGY ADKNEHPYVS IGRYMADKGY LKLGQTSMQG
 701  IKAYMRQNPQ RLAEVLGQNP SYIFFRELAG SSNDGPVGAL GTPLMGEYAG
 751  AVDRHYITLG APLFVATAHP VTRKALNRLI MAQDTGSAIK GAVRVDYFWG
 801  YGDEAGELAG KQKTTGYVWQ LLPNGMKPEY RP*
```

ΔG287–953

```
   1  ATGGCTAGCC CCGATGTTAA ATCGGCGGAC ACGCTGTCAA AACCGGCCGC
  51  TCCTGTTGTT GCTGAAAAAG AGACAGAGGT AAAAGAAGAT GCGCCACAGG
 101  CAGGTTCTCA AGGACAGGGC GCGCCATCCA CACAAGGCAG CCAAGATATG
 151  GCGGCAGTTT CGGCAGAAAA TACAGGCAAT GGCGGTGCGG CAACAACGGA
 201  CAAACCCAAA AATGAAGACG AGGGACCGCA AAATGATATG CCGCAAAATT
 251  CCGCCGAATC CGCAAATCAA ACAGGGAACA ACCAACCCGC CGATTCTTCA
 301  GATTCCGCCC CCGCGTCAAA CCCTGCACCT GCGAATGGCG GTAGCAATTT
 351  TGGAAGGGTT GATTTGGCTA ATGGCGTTTT GATTGATGGG CCGTCGCAAA
 401  ATATAACGTT GACCCACTGT AAAGGCGATT CTTGTAATGG TGATAATTTA
 451  TTGGATGAAG AAGCACCGTC AAAATCAGAA TTTGAAAATT TAAATGAGTC
 501  TGAACGAATT GAGAAATATA AGAAAGATGG GAAAAGCGAT AAATTTACTA
 551  ATTTGGTTGC GACAGCAGTT CAAGCTAATG GAACTAACAA ATATGTCATC
 601  ATTTATAAAG ACAAGTCCGC TTCATCTTCA TCTGCGCGAT TCAGGCGTTC
 651  TGCACGGTCG AGGAGGTCGC TTCCTGCCGA GATGCCGCTA ATCCCCGTCA
 701  ATCAGGCGGA TACGCTGATT GTCGATGGGG AAGCGGTCAG CCTGACGGGG
 751  CATTCCGGCA ATATCTTCGC GCCCGAAGGG AATTACCGGT ATCTGACTTA
```

```
 801  CGGGGCGGAA AAATTGCCCG GCGGATCGTA TGCCCTCCGT GTGCAAGGCG
 851  AACCGGCAAA AGGCGAAATG CTTGCTGGCA CGGCCGTGTA CAACGGCGAA
 901  GTGCTGCATT TTCATACGGA AAACGGCCGT CCGTACCCGA CTAGAGGCAG
 951  GTTTGCCGCA AAAGTCGATT TCGGCAGCAA ATCTGTGGAC GGCATTATCG
1001  ACAGCGGCGA TGATTTGCAT ATGGGTACGC AAAAATTCAA AGCCGCCATC
1051  GATGGAAACG GCTTTAAGGG GACTTGGACG GAAAATGGCG GCGGGGATGT
1101  TTCCGGAAGG TTTTACGGCC CGGCCGGCGA GGAAGTGGCG GGAAAATACA
1151  GCTATCGCCC GACAGATGCG GAAAAGGGCG GATTCGGCGT GTTTGCCGGC
1201  AAAAAAGAGC AGGATGGATC CGGAGGAGGA GGAGCCACCT ACAAAGTGGA
1251  CGAATATCAC GCCAACGCCC GTTTCGCCAT CGACCATTTC AACACCAGCA
1301  CCAACGTCGG CGGTTTTTAC GGTCTGACCG GTTCCGTCGA GTTCGACCAA
1351  GCAAAACGCG ACGGTAAAAT CGACATCACC ATCCCCGTTG CCAACCTGCA
1401  AAGCGGTTCG CAACACTTTA CCGACCACCT GAAATCAGCC GACATCTTCG
1451  ATGCCGCCCA ATATCCGGAC ATCCGCTTTG TTTCCACCAA ATTCAACTTC
1501  AACGGCAAAA AACTGGTTTC CGTTGACGGC AACCTGACCA TGCACGGCAA
1551  AACCGCCCCC GTCAAACTCA AAGCCGAAAA ATTCAACTGC TACCAAAGCC
1601  CGATGGCGAA AACCGAAGTT TGCGGCGGCG ACTTCAGCAC CACCATCGAC
1651  CGCACCAAAT GGGGCGTGGA CTACCTCGTT AACGTTGGTA TGACCAAAAG
1701  CGTCCGCATC GACATCCAAA TCGAGGCAGC CAAACAATAA CTCGAG

   1  MASPDVKSAD TLSKPAAPVV AEKETEVKED APQAGSQGQG APSTQGSQDM
  51  AAVSAENTGN GGAATTDKPK NEDEGPQNDM PQNSAESANQ TGNNQPADSS
 101  DSAPASNPAP ANGGSNFGRV DLANGVLIDG PSQNITLTHC KGDSCNGDNL
 151  LDEEAPSKSE FENLNESERI EKYKKDGKSD KFTNLVATAV QANGTNKYVI
 201  IYKDKSASSS SARFRRSARS RRSLPAEMPL IPVNQADTLI VDGEAVSLTG
 251  HSGNIFAPEG NYRYLTYGAE KLPGGSYALR VQGEPAKGEM LAGTAVYNGE
 301  VLHFHTENGR PYPTRGRFAA KVDFGSKSVD GIIDSGDDLH MGTQKFKAAI
 351  DGNGFKGTWT ENGGGDVSGR FYGPAGEEVA GKYSYRPTDA EKGGFGVFAG
 401  KKEQDGSGGG GATYKVDEYH ANARFAIDHF NTSTNVGGFY GLTGSVEFDQ
 451  AKRDGKIDIT IPVANLQSGS QHFTDHLKSA DIFDAAQYPD IRFVSTKFNF
 501  NGKKLVSVDG NLTMHGKTAP VKLKAEKFNC YQSPMAKTEV CGGDFSTTID
 551  RTKWGVDYLV NVGMTKSVRI DIQIEAAKQ*
```

ΔG287–961

```
   1  ATGGCTAGCC CCGATGTTAA ATCGGCGGAC ACGCTGTCAA AACCGGCCGC
  51  TCCTGTTGTT GCTGAAAAAG AGACAGAGGT AAAAGAAGAT GCGCCACAGG
 101  CAGGTTCTCA AGGACAGGGC GCGCCATCCA CACAAGGCAG CCAAGATATG
 151  GCGGCAGTTT CGGCAGAAAA TACAGGCAAT GGCGGTGCGG CAACAACGGA
 201  CAAACCCAAA AATGAAGACG AGGGACCGCA AAATGATATG CCGCAAAATT
 251  CCGCCGAATC CGCAAATCAA ACAGGGAACA ACCAACCCGC CGATTCTTCA
 301  GATTCCGCCC CCGCGTCAAA CCCTGCACCT GCGAATGGCG GTAGCAATTT
 351  TGGAAGGGTT GATTTGGCTA ATGGCGTTTT GATTGATGGG CCGTCGCAAA
 401  ATATAACGTT GACCCACTGT AAAGGCGATT CTTGTAATGG TGATAATTTA
 451  TTGGATGAAG AAGCACCGTC AAAATCAGAA TTTGAAAATT TAAATGAGTC
 501  TGAACGAATT GAGAAATATA AGAAAGATGG GAAAAGCGAT AAATTTACTA
 551  ATTTGGTTGC GACAGCAGTT CAAGCTAATG GAACTAACAA ATATGTCATC
 601  ATTTATAAAG ACAAGTCCGC TTCATCTTCA TCTGCGCGAT TCAGGCGTTC
 651  TGCACGGTCG AGGAGGTCGC TTCCTGCCGA GATGCCGCTA ATCCCCGTCA
 701  ATCAGGCGGA TACGCTGATT GTCGATGGGG AAGCGGTCAG CCTGACGGGG
 751  CATTCCGGCA ATATCTTCGC GCCCGAAGGG AATTACCGGT ATCTGACTTA
 801  CGGGGCGGAA AAATTGCCCG GCGGATCGTA TGCCCTCCGT GTGCAAGGCG
 851  AACCGGCAAA AGGCGAAATG CTTGCTGGCA CGGCCGTGTA CAACGGCGAA
 901  GTGCTGCATT TTCATACGGA AAACGGCCGT CCGTACCCGA CTAGAGGCAG
 951  GTTTGCCGCA AAAGTCGATT TCGGCAGCAA ATCTGTGGAC GGCATTATCG
1001  ACAGCGGCGA TGATTTGCAT ATGGGTACGC AAAAATTCAA AGCCGCCATC
1051  GATGGAAACG GCTTTAAGGG GACTTGGACG GAAAATGGCG GCGGGGATGT
1101  TTCCGGAAGG TTTTACGGCC CGGCCGGCGA GGAAGTGGCG GGAAAATACA
1151  GCTATCGCCC GACAGATGCG GAAAAGGGCG GATTCGGCGT GTTTGCCGGC
1201  AAAAAAGAGC AGGATGGATC CGGAGGAGGA GGAGCCACAA ACGACGACGA
1251  TGTTAAAAAA GCTGCCACTG TGGCCATTGC TGCTGCCTAC AACAATGGCC
1301  AAGAAATCAA CGGTTTCAAA GCTGGAGAGA CCATCTACGA CATTGATGAA
1351  GACGGCACAA TTACCAAAAA AGACGCAACT GCAGCCGATG TTGAAGCCGA
1401  CGACTTTAAA GGTCTGGGTC TGAAAAAAGT CGTGACTAAC CTGACCAAAA
1451  CCGTCAATGA AAACAAACAA AACGTCGATG CCAAAGTAAA AGCTGCAGAA
1501  TCTGAAATAG AAAAGTTAAC AACCAAGTTA GCAGACACTG ATGCCGCTTT
1551  AGCAGATACT GATGCCGCTC TGGATGCAAC CACCAACGCC TTGAATAAAT
```

```
1601   TGGGAGAAAA TATAACGACA TTTGCTGAAG AGACTAAGAC AAATATCGTA
1651   AAAATTGATG AAAAATTAGA AGCCGTGGCT GATACCGTCG ACAAGCATGC
1701   CGAAGCATTC AACGATATCG CCGATTCATT GGATGAAACC AACACTAAGG
1751   CAGACGAAGC CGTCAAAACC GCCAATGAAG CCAAACAGAC GGCCGAAGAA
1801   ACCAAACAAA ACGTCGATGC CAAAGTAAAA GCTGCAGAAA CTGCAGCAGG
1851   CAAAGCCGAA GCTGCCGCTG GCACAGCTAA TACTGCAGCC GACAAGGCCG
1901   AAGCTGTCGC TGCAAAAGTT ACCGACATCA AAGCTGATAT CGCTACGAAC
1951   AAAGATAATA TTGCTAAAAA AGCAAACAGT GCCGACGTGT ACACCAGAGA
2001   AGAGTCTGAC AGCAAATTTG TCAGAATTGA TGGTCTGAAC GCTACTACCG
2051   AAAAATTGGA CACACGCTTG GCTTCTGCTG AAAAATCCAT TGCCGATCAC
2101   GATACTCGCC TGAACGGTTT GGATAAAACA GTGTCAGACC TGCGCAAAGA
2151   AACCCGCCAA GGCCTTGCAG AACAAGCCGC GCTCTCCGGT CTGTTCCAAC
2201   CTTACAACGT GGGTCGGTTC AATGTAACGG CTGCAGTCGG CGGCTACAAA
2251   TCCGAATCGG CAGTCGCCAT CGGTACCGGC TTCCGCTTTA CCGAAAACTT
2301   TGCCGCCAAA GCAGGCGTGG CAGTCGGCAC TTCGTCCGGT TCTTCCGCAG
2351   CCTACCATGT CGGCGTCAAT TACGAGTGGT AACTCGAG

1      MASPDVKSAD TLSKPAAPVV AEKETEVKED APQAGSQGQG APSTQGSQDM
51     AAVSAENTGN GGAATTDKPK NEDEGPQNDM PQNSAESANQ TGNNQPADSS
101    DSAPASNPAP ANGGSNFGRV DLANGVLIDG PSQNITLTHC KGDSCNGDNL
151    LDEEAPSKSE FENLNESERI EKYKKDGKSD KFTNLVATAV QANGTNKYVI
201    IYKDKSASSS SARFRRSARS RRSLPAEMPL IPVNQADTLI VDGEAVSLTG
251    HSGNIFAPEG NYRYLTYGAE KLPGGSYALR VQGEPAKGEM LAGTAVYNGE
301    VLHFHTENGR PYPTRGRFAA KVDFGSKSVD GIIDSGDDLH MGTQKFKAAI
351    DGNGFKGTWT ENGGGDVSGR FYGPAGEEVA GKYSYRPTDA EKGGFGVFAG
401    KKEQDGSGGG GATNDDDVKK AATVAIAAAY NNGQEINGFK AGETIYDIDE
451    DGTITKKDAT AADVEADDFK GLGLKKVVTN LTKTVNENKQ NVDAKVKAAE
501    SEIEKLTTKL ADTDAALADT DAALDATTNA LNKLGENITT FAEETKTNIV
551    KIDEKLEAVA DTVDKHAEAF NDIADSLDET NTKADEAVKT ANEAKQTAEE
601    TKQNVDAKVK AAETAAGKAE AAAGTANTAA DKAEAVAAKV TDIKADIATN
651    KDNIAKKANS ADVYTREESD SKFVRIDGLN ATTEKLDTRL ASAEKSIADH
701    DTRLNGLDKT VSDLRKETRQ GLAEQAALSG LFQPYNVGRF NVTAAVGGYK
751    SESAVAIGTG FRFTENFAAK AGVAVGTSSG SSAAYHVGVN YEW*
```

	ELISA	Bactericidal
ΔG287-953-His	3834	65536
ΔG287-961-His	108627	65536

The bactericidal efficacy (homologous strain) of antibodies raised against the hybrid proteins was compared with antibodies raised against simple mixtures of the component antigens (using 287-GST) for 919 and ORF46.1:

	Mixture with 287	Hybrid with ΔG287
919	32000	128000
ORF46.1	128	16000

Data for bactericidal activity against heterologous MenB strains and against serotypes A and C were also obtained:

	919		ORF46.1	
Strain	*Mixture*	*Hybrid*	*Mixture*	*Hybrid*
NGH38	1024	32000	-	16384
MC58	512	8192	-	512
BZ232	512	512	-	-
MenA (F6124)	512	32000	-	8192
MenC (C11)	>2048	>2048	-	-
MenC (BZ133)	>4096	64000	-	8192

The hybrid proteins with ΔG287 at the N-terminus are therefore immunologically superior to simple mixtures, with ΔG287-ORF46.1 being particularly effective, even against heterologous strains. ΔG287-ORF46.1K may be expressed in pET-24b.

The same hybrid proteins were made using New Zealand strain 394/98 rather than 2996:

ΔG287NZ-919

```
   1  ATGGCTAGCC CCGATGTCAA GTCGGCGGAC ACGCTGTCAA AACCTGCCGC
  51  CCCTGTTGTT TCTGAAAAAG AGACAGAGGC AAAGGAAGAT GCGCCACAGG
 101  CAGGTTCTCA AGGACAGGGC GCGCCATCCG CACAAGGCGG TCAAGATATG
 151  GCGGCGGTTT CGGAAGAAAA TACAGGCAAT GGCGGTGCGG CAGCAACGGA
 201  CAAACCCAAA AATGAAGACG AGGGGGCGCA AAATGATATG CCGCAAAATG
 251  CCGCCGATAC AGATAGTTTG ACACCGAATC ACACCCCGGC TTCGAATATG
 301  CCGGCCGGAA ATATGGAAAA CCAAGCACCG GATGCCGGGG AATCGGAGCA
 351  GCCGGCAAAC CAACCGGATA TGGCAAATAC GGCGGACGGA ATGCAGGGTG
 401  ACGATCCGTC GGCAGGCGGG GAAAATGCCG GCAATACGGC TGCCCAAGGT
 451  ACAAATCAAG CCGAAAACAA TCAAACCGCC GGTTCTCAAA ATCCTGCCTC
 501  TTCAACCAAT CCTAGCGCCA CGAATAGCGG TGGTGATTTT GGAAGGACGA
 551  ACGTGGGCAA TTCTGTTGTG ATTGACGGGC CGTCGCAAAA TATAACGTTG
 601  ACCCACTGTA AAGGCGATTC TTGTAGTGGC AATAATTTCT TGGATGAAGA
 651  AGTACAGCTA AAATCAGAAT TTGAAAAATT AAGTGATGCA GACAAAATAA
 701  GTAATTACAA GAAAGATGGG AAGAATGACG GGAAGAATGA TAAATTTGTC
 751  GGTTTGGTTG CCGATAGTGT GCAGATGAAG GGAATCAATC AATATATTAT
 801  CTTTTATAAA CCTAAACCCA CTTCATTTGC GCGATTTAGG CGTTCTGCAC
 851  GGTCGAGGCG GTCGCTTCCG GCCGAGATGC CGCTGATTCC CGTCAATCAG
 901  GCGGATACGC TGATTGTCGA TGGGGAAGCG GTCAGCCTGA CGGGGCATTC
 951  CGGCAATATC TTCGCGCCCG AAGGGAATTA CCGGTATCTG ACTTACGGGG
1001  CGGAAAAATT GCCCGGCGGA TCGTATGCCC TCCGTGTTCA AGGCGAACCT
1051  TCAAAAGGCG AAATGCTCGC GGGCACGGCA GTGTACAACG GCGAAGTGCT
1101  GCATTTTCAT ACGGAAAACG GCCGTCCGTC CCCGTCCAGA GGCAGGTTTG
1151  CCGCAAAAGT CGATTTCGGC AGCAAATCTG TGGACGGCAT TATCGACAGC
1201  GGCGATGGTT TGCATATGGG TACGCAAAAA TTCAAAGCCG CCATCGATGG
1251  AAACGGCTTT AAGGGGACTT GGACGGAAAA TGGCGGCGGG GATGTTTCCG
1301  GAAAGTTTTA CGGCCCGGCC GGCGAGGAAG TGGCGGGAAA ATACAGCTAT
1351  CGCCCAACAG ATGCGGAAAA GGGCGGATTC GGCGTGTTTG CCGGCAAAAA
1401  AGAGCAGGAT GGATCCGGAG GAGGAGGATG CCAAAGCAAG AGCATCCAAA
1451  CCTTTCCGCA ACCCGACACA TCCGTCATCA ACGGCCCGGA CCGGCCGGTC
1501  GGCATCCCCG ACCCCGCCGG AACGACGGTC GGCGGCGGCG GGGCCGTCTA
1551  TACCGTTGTA CCGCACCTGT CCCTGCCCCA CTGGGCGGCG CAGGATTTCG
1601  CCAAAAGCCT GCAATCCTTC CGCCTCGGCT GCGCCAATTT GAAAAACCGC
1651  CAAGGCTGGC AGGATGTGTG CGCCCAAGCC TTTCAAACCC CCGTCCATTC
1701  CTTTCAGGCA AAACAGTTTT TTGAACGCTA TTTCACGCCG TGGCAGGTTG
1751  CAGGCAACGG AAGCCTTGCC GGTACGGTTA CCGGCTATTA CGAGCCGGTG
1801  CTGAAGGGCG ACGACAGGCG GACGGCACAA GCCCGCTTCC CGATTTACGG
1851  TATTCCCGAC GATTTTATCT CCGTCCCCCT GCCTGCCGGT TTGCGGAGCG
1901  GAAAAGCCCT TGTCCGCATC AGGCAGACGG GAAAAAACAG CGGCACAATC
1951  GACAATACCG GCGGCACACA TACCGCCGAC CTCTCCCGAT TCCCCATCAC
2001  CGCGCGCACA ACGGCAATCA AAGGCAGGTT TGAAGGAAGC CGCTTCCTCC
```

```
2051  CCTACCACAC GCGCAACCAA ATCAACGGCG GCGCGCTTGA CGGCAAAGCC
2101  CCGATACTCG GTTACGCCGA AGACCCCGTC GAACTTTTTT TTATGCACAT
2151  CCAAGGCTCG GGCCGTCTGA AAACCCCGTC CGGCAAATAC ATCCGCATCG
2201  GCTATGCCGA CAAAAACGAA CATCCCTACG TTTCCATCGG ACGCTATATG
2251  GCGGACAAAG GCTACCTCAA GCTCGGGCAG ACCTCGATGC AGGGCATCAA
2301  AGCCTATATG CGGCAAAATC CGCAACGCCT CGCCGAAGTT TTGGGTCAAA
2351  ACCCCAGCTA TATCTTTTTC CGCGAGCTTG CCGGAAGCAG CAATGACGGT
2401  CCCGTCGGCG CACTGGGCAC GCCGTTGATG GGGGAATATG CCGGCGCAGT
2451  CGACCGGCAC TACATTACCT TGGGCGCGCC CTTATTTGTC GCCACCGCCC
2501  ATCCGGTTAC CCGCAAAGCC CTCAACCGCC TGATTATGGC GCAGGATACC
2551  GGCAGCGCGA TTAAAGGCGC GGTGCGCGTG GATTATTTTT GGGGATACGG
2601  CGACGAAGCC GGCGAACTTG CCGGCAAACA GAAAACCACG GGTTACGTCT
2651  GGCAGCTCCT ACCCAACGGT ATGAAGCCCG AATACCGCCC GTAAAAGCTT

   1  MASPDVKSAD TLSKPAAPVV SEKETEAKED APQAGSQGQG APSAQGGQDM
  51  AAVSEENTGN GGAAATDKPK NEDEGAQNDM PQNAADTDSL TPNHTPASNM
 101  PAGNMENQAP DAGESEQPAN QPDMANTADG MQGDDPSAGG ENAGNTAAQG
 151  TNQAENNQTA GSQNPASSTN PSATNSGGDF GRTNVGNSVV IDGPSQNITL
 201  THCKGDSCSG NNFLDEEVQL KSEFEKLSDA DKISNYKKDG KNDGKNDKFV
 251  GLVADSVQMK GINQYIIFYK PKPTSFARFR RSARSRRSLP AEMPLIPVNQ
 301  ADTLIVDGEA VSLTGHSGNI FAPEGNYRYL TYGAEKLPGG SYALRVQGEP
 351  SKGEMLAGTA VYNGEVLHFH TENGRPSPSR GRFAAKVDFG SKSVDGIIDS
 401  GDGLHMGTQK FKAAIDGNGF KGTWTENGGG DVSGKFYGPA GEEVAGKYSY
 451  RPTDAEKGGF GVFAGKKEQD GSGGGGCQSK SIQTFPQPDT SVINGPDRPV
 501  GIPDPAGTTV GGGGAVYTVV PHLSLPHWAA QDFAKSLQSF RLGCANLKNR
 551  QGWQDVCAQA FQTPVHSFQA KQFFERYFTP WQVAGNGSLA GTVTGYYEPV
 601  LKGDDRRTAQ ARFPIYGIPD DFISVPLPAG LRSGKALVRI RQTGKNSGTI
 651  DNTGGTHTAD LSRFPITART TAIKGRFEGS RFLPYHTRNQ INGGALDGKA
 701  PILGYAEDPV ELFFMHIQGS GRLKTPSGKY IRIGYADKNE HPYVSIGRYM
 751  ADKGYLKLGQ TSMQGIKAYM RQNPQRLAEV LGQNPSYIFF RELAGSSNDG
 801  PVGALGTPLM GEYAGAVDRH YITLGAPLFV ATAHPVTRKA LNRLIMAQDT
 851  GSAIKGAVRV DYFWGYGDEA GELAGKQKTT GYVWQLLPNG MKPEYRP*
```

ΔG287NZ-953

```
   1  ATGGCTAGCC CCGATGTCAA GTCGGCGGAC ACGCTGTCAA AACCTGCCGC
  51  CCCTGTTGTT TCTGAAAAAG AGACAGAGGC AAAGGAAGAT GCGCCACAGG
 101  CAGGTTCTCA AGGACAGGGC GCGCCATCCG CACAAGGCGG TCAAGATATG
 151  GCGGCGGTTT CGGAAGAAAA TACAGGCAAT GGCGGTGCGG CAGCAACGGA
 201  CAAACCCAAA AATGAAGACG AGGGGGCGCA AAATGATATG CCGCAAAATG
 251  CCGCCGATAC AGATAGTTTG ACACCGAATC ACACCCCGGC TTCGAATATG
 301  CCGGCCGGAA ATATGGAAAA CCAAGCACCG GATGCCGGGG AATCGGAGCA
 351  GCCGGCAAAC CAACCGGATA TGGCAAATAC GGCGGACGGA ATGCAGGGTG
 401  ACGATCCGTC GGCAGGCGGG GAAAATGCCG GCAATACGGC TGCCCAAGGT
 451  ACAAATCAAG CCGAAAACAA TCAAACCGCC GGTTCTCAAA ATCCTGCCTC
 501  TTCAACCAAT CCTAGCGCCA CGAATAGCGG TGGTGATTTT GGAAGGACGA
 551  ACGTGGGCAA TTCTGTTGTG ATTGACGGGC CGTCGCAAAA TATAACGTTG
 601  ACCCACTGTA AAGGCGATTC TTGTAGTGGC AATAATTTCT TGGATGAAGA
 651  AGTACAGCTA AAATCAGAAT TTGAAAAATT AAGTGATGCA GACAAAATAA
 701  GTAATTACAA GAAAGATGGG AAGAATGACG GGAAGAATGA TAAATTTGTC
 751  GGTTTGGTTG CCGATAGTGT GCAGATGAAG GGAATCAATC AATATATTAT
 801  CTTTTATAAA CCTAAACCCA CTTCATTTGC GCGATTTAGG CGTTCTGCAC
 851  GGTCGAGGCG GTCGCTTCCG GCCGAGATGC CGCTGATTCC CGTCAATCAG
 901  GCGGATACGC TGATTGTCGA TGGGGAAGCG GTCAGCCTGA CGGGGCATTC
 951  CGGCAATATC TTCGCGCCCG AAGGGAATTA CCGGTATCTG ACTTACGGGG
1001  CGGAAAAATT GCCCGGCGGA TCGTATGCCC TCCGTGTTCA AGGCGAACCT
1051  TCAAAAGGCG AAATGCTCGC GGGCACGGCA GTGTACAACG GCGAAGTGCT
1101  GCATTTTCAT ACGGAAAACG GCCGTCCGTC CCCGTCCAGA GGCAGGTTTG
1151  CCGCAAAAGT CGATTTCGGC AGCAAATCTG TGGACGGCAT TATCGACAGC
1201  GGCGATGGTT TGCATATGGG TACGCAAAAA TTCAAAGCCG CCATCGATGG
1251  AAACGGCTTT AAGGGGACTT GGACGGAAAA TGGCGGCGGG GATGTTTCCG
1301  GAAAGTTTTA CGGCCCGGCC GGCGAGGAAG TGGCGGGAAA ATACAGCTAT
1351  CGCCCAACAG ATGCGGAAAA GGGCGGATTC GGCGTGTTTG CCGGCAAAAA
1401  AGAGCAGGAT GGATCCGGAG GAGGAGGAGC CACCTACAAA GTGGACGAAT
1451  ATCACGCCAA CGCCCGTTTC GCCATCGACC ATTTCAACAC CAGCACCAAC
1501  GTCGGCGGTT TTTACGGTCT GACCGGTTCC GTCGAGTTCG ACCAAGCAAA
1551  ACGCGACGGT AAAATCGACA TCACCATCCC CGTTGCCAAC CTGCAAAGCG
```

```
1601  GTTCGCAACA CTTTACCGAC CACCTGAAAT CAGCCGACAT CTTCGATGCC
1651  GCCCAATATC CGGACATCCG CTTTGTTTCC ACCAAATTCA ACTTCAACGG
1701  CAAAAAACTG GTTTCCGTTG ACGGCAACCT GACCATGCAC GGCAAAACCG
1751  CCCCCGTCAA ACTCAAAGCC GAAAAATTCA ACTGCTACCA AAGCCCGATG
1801  GCGAAAACCG AAGTTTGCGG CGGCGACTTC AGCACCACCA TCGACCGCAC
1851  CAAATGGGGC GTGGACTACC TCGTTAACGT TGGTATGACC AAAAGCGTCC
1901  GCATCGACAT CCAAATCGAG GCAGCCAAAC AATAAAAGCT T

   1  MASPDVKSAD TLSKPAAPVV SEKETEAKED APQAGSQGQG APSAQGGQDM
  51  AAVSEENTGN GGAAATDKPK NEDEGAQNDM PQNAADTDSL TPNHTPASNM
 101  PAGNMENQAP DAGESEQPAN QPDMANTADG MQGDDPSAGG ENAGNTAAQG
 151  TNQAENNQTA GSQNPASSTN PSATNSGGDF GRTNVGNSVV IDGPSQNITL
 201  THCKGDSCSG NNFLDEEVQL KSEFEKLSDA DKISNYKKDG KNDGKNDKFV
 251  GLVADSVQMK GINQYIIFYK PKPTSFARFR RSARSRRSLP AEMPLIPVNQ
 301  ADTLIVDGEA VSLTGHSGNI FAPEGNYRYL TYGAEKLPGG SYALRVQGEP
 351  SKGEMLAGTA VYNGEVLHFH TENGRPSPSR GRFAAKVDFG SKSVDGIIDS
 401  GDGLHMGTQK FKAAIDGNGF KGTWTENGGG DVSGKFYGPA GEEVAGKYSY
 451  RPTDAEKGGF GVFAGKKEQD GSGGGGATYK VDEYHANARF AIDHFNTSTN
 501  VGGFYGLTGS VEFDQAKRDG KIDITIPVAN LQSGSQHFTD HLKSADIFDA
 551  AQYPDIRFVS TKFNFNGKKL VSVDGNLTMH GKTAPVKLKA EKFNCYQSPM
 601  AKTEVCGGDF STTIDRTKWG VDYLVNVGMT KSVRIDIQIE AAKQ*
```

ΔG287NZ-961

```
   1  ATGGCTAGCC CCGATGTCAA GTCGGCGGAC ACGCTGTCAA AACCTGCCGC
  51  CCCTGTTGTT TCTGAAAAAG AGACAGAGGC AAAGGAAGAT GCGCCACAGG
 101  CAGGTTCTCA AGGACAGGGC GCGCCATCCG CACAAGGCGG TCAAGATATG
 151  GCGGCGGTTT CGGAAGAAAA TACAGGCAAT GGCGGTGCGG CAGCAACGGA
 201  CAAACCCAAA AATGAAGACG AGGGGGCGCA AAATGATATG CCGCAAAATG
 251  CCGCCGATAC AGATAGTTTG ACACCGAATC ACACCCCGGC TTCGAATATG
 301  CCGGCCGGAA ATATGGAAAA CCAAGCACCG GATGCCGGGG AATCGGAGCA
 351  GCCGGCAAAC CAACCGGATA TGGCAAATAC GGCGGACGGA ATGCAGGGTG
 401  ACGATCCGTC GGCAGGCGGG GAAAATGCCG GCAATACGGC TGCCCAAGGT
 451  ACAAATCAAG CCGAAAACAA TCAAACCGCC GGTTCTCAAA ATCCTGCCTC
 501  TTCAACCAAT CCTAGCGCCA CGAATAGCGG TGGTGATTTT GGAAGGACGA
 551  ACGTGGGCAA TTCTGTTGTG ATTGACGGGC CGTCGCAAAA TATAACGTTG
 601  ACCCACTGTA AAGGCGATTC TTGTAGTGGC AATAATTTCT TGGATGAAGA
 651  AGTACAGCTA AAATCAGAAT TTGAAAAATT AAGTGATGCA GACAAAATAA
 701  GTAATTACAA GAAAGATGGG AAGAATGACG GGAAGAATGA TAAATTTGTC
 751  GGTTTGGTTG CCGATAGTGT GCAGATGAAG GGAATCAATC AATATATTAT
 801  CTTTTATAAA CCTAAACCCA CTTCATTTGC GCGATTTAGG CGTTCTGCAC
 851  GGTCGAGGCG GTCGCTTCCG GCCGAGATGC CGCTGATTCC CGTCAATCAG
 901  GCGGATACGC TGATTGTCGA TGGGGAAGCG GTCAGCCTGA CGGGGCATTC
 951  CGGCAATATC TTCGCGCCCG AAGGGAATTA CCGGTATCTG ACTTACGGGG
1001  CGGAAAAATT GCCCGGCGGA TCGTATGCCC TCCGTGTTCA AGGCGAACCT
1051  TCAAAAGGCG AAATGCTCGC GGGCACGGCA GTGTACAACG GCGAAGTGCT
1101  GCATTTTCAT ACGGAAAACG GCCGTCCGTC CCCGTCCAGA GGCAGGTTTG
1151  CCGCAAAAGT CGATTTCGGC AGCAAATCTG TGGACGGCAT TATCGACAGC
1201  GGCGATGGTT TGCATATGGG TACGCAAAAA TTCAAAGCCG CCATCGATGG
1251  AAACGGCTTT AAGGGGACTT GGACGGAAAA TGGCGGCGGG GATGTTTCCG
1301  GAAAGTTTTA CGGCCCGGCC GGCGAGGAAG TGGCGGGAAA ATACAGCTAT
1351  CGCCCAACAG ATGCGGAAAA GGGCGGATTC GGCGTGTTTG CCGGCAAAAA
1401  AGAGCAGGAT GGATCCGGAG GAGGAGGAGC CACAAACGAC GACGATGTTA
1451  AAAAAGCTGC CACTGTGGCC ATTGCTGCTG CCTACAACAA TGGCCAAGAA
1501  ATCAACGGTT TCAAAGCTGG AGAGACCATC TACGACATTG ATGAAGACGG
1551  CACAATTACC AAAAAAGACG CAACTGCAGC CGATGTTGAA GCCGACGACT
1601  TTAAAGGTCT GGGTCTGAAA AAAGTCGTGA CTAACCTGAC CAAAACCGTC
1651  AATGAAAACA AACAAAACGT CGATGCCAAA GTAAAAGCTG CAGAATCTGA
1701  AATAGAAAAG TTAACAACCA AGTTAGCAGA CACTGATGCC GCTTTAGCAG
1751  ATACTGATGC CGCTCTGGAT GCAACCACCA ACGCCTTGAA TAAATTGGGA
1801  GAAAATATAA CGACATTTGC TGAAGAGACT AAGACAAATA TCGTAAAAAT
1851  TGATGAAAAA TTAGAAGCCG TGGCTGATAC CGTCGACAAG CATGCCGAAG
1901  CATTCAACGA TATCGCCGAT TCATTGGATG AAACCAACAC TAAGGCAGAC
1951  GAAGCCGTCA AAACCGCCAA TGAAGCCAAA CAGACGGCCG AAGAAACCAA
2001  ACAAAACGTC GATGCCAAAG TAAAAGCTGC AGAAACTGCA GCAGGCAAAG
2051  CCGAAGCTGC CGCTGGCACA GCTAATACTG CAGCCGACAA GGCCGAAGCT
2101  GTCGCTGCAA AAGTTACCGA CATCAAAGCT GATATCGCTA CGAACAAAGA
```

```
2151  TAATATTGCT AAAAAAGCAA ACAGTGCCGA CGTGTACACC AGAGAAGAGT
2201  CTGACAGCAA ATTTGTCAGA ATTGATGGTC TGAACGCTAC TACCGAAAAA
2251  TTGGACACAC GCTTGGCTTC TGCTGAAAAA TCCATTGCCG ATCACGATAC
2301  TCGCCTGAAC GGTTTGGATA AAACAGTGTC AGACCTGCGC AAAGAAACCC
2351  GCCAAGGCCT TGCAGAACAA GCCGCGCTCT CCGGTCTGTT CCAACCTTAC
2401  AACGTGGGTC GGTTCAATGT AACGGCTGCA GTCGGCGGCT ACAAATCCGA
2451  ATCGGCAGTC GCCATCGGTA CCGGCTTCCG CTTTACCGAA AACTTTGCCG
2501  CCAAAGCAGG CGTGGCAGTC GGCACTTCGT CCGGTTCTTC CGCAGCCTAC
2551  CATGTCGGCG TCAATTACGA GTGGTAAAAG CTT

1     MASPDVKSAD TLSKPAAPVV SEKETEAKED APQAGSQGQG APSAQGGQDM
51    AAVSEENTGN GGAAATDKPK NEDEGAQNDM PQNAADTDSL TPNHTPASNM
101   PAGNMENQAP DAGESEQPAN QPDMANTADG MQGDDPSAGG ENAGNTAAQG
151   TNQAENNQTA GSQNPASSTN PSATNSGGDF GRTNVGNSVV IDGPSQNITL
201   THCKGDSCSG NNFLDEEVQL KSEFEKLSDA DKISNYKKDG KNDGKNDKFV
251   GLVADSVQMK GINQYIIFYK PKPTSFARFR RSARSRRSLP AEMPLIPVNQ
301   ADTLIVDGEA VSLTGHSGNI FAPEGNYRYL TYGAEKLPGG SYALRVQGEP
351   SKGEMLAGTA VYNGEVLHFH TENGRPSPSR GRFAAKVDFG SKSVDGIIDS
401   GDGLHMGTQK FKAAIDGNGF KGTWTENGGG DVSGKFYGPA GEEVAGKYSY
451   RPTDAEKGGF GVFAGKKEQD GSGGGGATND DDVKKAATVA IAAAYNNGQE
501   INGFKAGETI YDIDEDGTIT KKDATAADVE ADDFKGLGLK KVVTNLTKTV
551   NENKQNVDAK VKAAESEIEK LTTKLADTDA ALADTDAALD ATTNALNKLG
601   ENITTFAEET KTNIVKIDEK LEAVADTVDK HAEAFNDIAD SLDETNTKAD
651   EAVKTANEAK QTAEETKQNV DAKVKAAETA AGKAEAAAGT ANTAADKAEA
701   VAAKVTDIKA DIATNKDNIA KKANSADVYT REESDSKFVR IDGLNATTEK
751   LDTRLASAEK SIADHDTRLN GLDKTVSDLR KETRQGLAEQ AALSGLFQPY
801   NVGRFNVTAA VGGYKSESAV AIGTGFRFTE NFAAKAGVAV GTSSGSSAAY
851   HVGVNYEW*
```

ΔG983 and hybrids

Bactericidal titres generated in response to ΔG983 (His-fusion) were measured against various strains, including the homologous 2996 strain:

	2996	NGH38	BZ133
ΔG983	512	128	128

ΔG983 was also expressed as a hybrid, with ORF46.1, 741, 961 or 961c at its C-terminus:

ΔG983-ORF46.1

```
1     ATGACTTCTG CGCCCGACTT CAATGCAGGC GGTACCGGTA TCGGCAGCAA
51    CAGCAGAGCA ACAACAGCGA AATCAGCAGC AGTATCTTAC GCCGGTATCA
101   AGAACGAAAT GTGCAAAGAC AGAAGCATGC TCTGTGCCGG TCGGGATGAC
151   GTTGCGGTTA CAGACAGGGA TGCCAAAATC AATGCCCCCC CCCCGAATCT
201   GCATACCGGA GACTTTCCAA ACCCAAATGA CGCATACAAG AATTTGATCA
251   ACCTCAAACC TGCAATTGAA GCAGGCTATA CAGGACGCGG GGTAGAGGTA
301   GGTATCGTCG ACACAGGCGA ATCCGTCGGC AGCATATCCT TTCCCGAACT
351   GTATGGCAGA AAAGAACACG GCTATAACGA AAATTACAAA AACTATACGG
401   CGTATATGCG GAAGGAAGCG CCTGAAGACG GAGGCGGTAA AGACATTGAA
451   GCTTCTTTCG ACGATGAGGC CGTTATAGAG ACTGAAGCAA AGCCGACGGA
501   TATCCGCCAC GTAAAAGAAA TCGGACACAT CGATTTGGTC TCCCATATTA
551   TTGGCGGGCG TTCCGTGGAC GGCAGACCTG CAGGCGGTAT TGCGCCCGAT
601   GCGACGCTAC ACATAATGAA TACGAATGAT GAAACCAAGA ACGAAATGAT
651   GGTTGCAGCC ATCCGCAATG CATGGGTCAA GCTGGGCGAA CGTGGCGTGC
701   GCATCGTCAA TAACAGTTTT GGAACAACAT CGAGGGCAGG CACTGCCGAC
751   CTTTTCCAAA TAGCCAATTC GGAGGAGCAG TACCGCCAAG CGTTGCTCGA
801   CTATTCCGGC GGTGATAAAA CAGACGAGGG TATCCGCCTG ATGCAACAGA
851   GCGATTACGG CAACCTGTCC TACCACATCC GTAATAAAAA CATGCTTTTC
901   ATCTTTTCGA CAGGCAATGA CGCACAAGCT CAGCCCAACA CATATGCCCT
951   ATTGCCATTT TATGAAAAAG ACGCTCAAAA AGGCATTATC ACAGTCGCAG
1001  GCGTAGACCG CAGTGGAGAA AAGTTCAAAC GGGAAATGTA TGGAGAACCG
1051  GGTACAGAAC CGCTTGAGTA TGGCTCCAAC CATTGCGGAA TTACTGCCAT
1101  GTGGTGCCTG TCGGCACCCT ATGAAGCAAG CGTCCGTTTC ACCCGTACAA
```

```
1151  ACCCGATTCA AATTGCCGGA ACATCCTTTT CCGCACCCAT CGTAACCGGC
1201  ACGGCGGCTC TGCTGCTGCA GAAATACCCG TGGATGAGCA ACGACAACCT
1251  GCGTACCACG TTGCTGACGA CGGCTCAGGA CATCGGTGCA GTCGGCGTGG
1301  ACAGCAAGTT CGGCTGGGGA CTGCTGGATG CGGGTAAGGC CATGAACGGA
1351  CCCGCGTCCT TTCCGTTCGG CGACTTTACC GCCGATACGA AAGGTACATC
1401  CGATATTGCC TACTCCTTCC GTAACGACAT TTCAGGCACG GGCGGCCTGA
1451  TCAAAAAAGG CGGCAGCCAA CTGCAACTGC ACGGCAACAA CACCTATACG
1501  GGCAAAACCA TTATCGAAGG CGGTTCGCTG GTGTTGTACG GCAACAACAA
1551  ATCGGATATG CGCGTCGAAA CCAAAGGTGC GCTGATTTAT AACGGGGCGG
1601  CATCCGGCGG CAGCCTGAAC AGCGACGGCA TTGTCTATCT GGCAGATACC
1651  GACCAATCCG GCGCAAACGA AACCGTACAC ATCAAAGGCA GTCTGCAGCT
1701  GGACGGCAAA GGTACGCTGT ACACACGTTT GGGCAAACTG CTGAAAGTGG
1751  ACGGTACGGC GATTATCGGC GGCAAGCTGT ACATGTCGGC ACGCGGCAAG
1801  GGGGCAGGCT ATCTCAACAG TACCGGACGA CGTGTTCCCT TCCTGAGTGC
1851  CGCCAAAATC GGGCAGGATT ATTCTTTCTT CACAAACATC GAAACCGACG
1901  GCGGCCTGCT GGCTTCCCTC GACAGCGTCG AAAAAACAGC GGGCAGTGAA
1951  GGCGACACGC TGTCCTATTA TGTCCGTCGC GGCAATGCGG CACGGACTGC
2001  TTCGGCAGCG GCACATTCCG CGCCCGCCGG TCTGAAACAC GCCGTAGAAC
2051  AGGGCGGCAG CAATCTGGAA AACCTGATGG TCGAACTGGA TGCCTCCGAA
2101  TCATCCGCAA CACCCGAGAC GGTTGAAACT GCGGCAGCCG ACCGCACAGA
2151  TATGCCGGGC ATCCGCCCCT ACGGCGCAAC TTTCCGCGCA GCGGCAGCCG
2201  TACAGCATGC GAATGCCGCC GACGGTGTAC GCATCTTCAA CAGTCTCGCC
2251  GCTACCGTCT ATGCCGACAG TACCGCCGCC CATGCCGATA TGCAGGGACG
2301  CCGCCTGAAA GCCGTATCGG ACGGGTTGGA CCACAACGGC ACGGGTCTGC
2351  GCGTCATCGC GCAAACCCAA CAGGACGGTG GAACGTGGGA ACAGGGCGGT
2401  GTTGAAGGCA AAATGCGCGG CAGTACCCAA ACCGTCGGCA TTGCCGCGAA
2451  AACCGGCGAA AATACGACAG CAGCCGCCAC ACTGGGCATG GGACGCAGCA
2501  CATGGAGCGA AAACAGTGCA AATGCAAAAA CCGACAGCAT TAGTCTGTTT
2551  GCAGGCATAC GGCACGATGC GGGCGATATC GGCTATCTCA AAGGCCTGTT
2601  CTCCTACGGA CGCTACAAAA ACAGCATCAG CCGCAGCACC GGTGCGGACG
2651  AACATGCGGA AGGCAGCGTC AACGGCACGC TGATGCAGCT GGGCGCACTG
2701  GGCGGTGTCA ACGTTCCGTT TGCCGCAACG GGAGATTTGA CGGTCGAAGG
2751  CGGTCTGCGC TACGACCTGC TCAAACAGGA TGCATTCGCC GAAAAAGGCA
2801  GTGCTTTGGG CTGGAGCGGC AACAGCCTCA CTGAAGGCAC GCTGGTCGGA
2851  CTCGCGGGTC TGAAGCTGTC GCAACCCTTG AGCGATAAAG CCGTCCTGTT
2901  TGCAACGGCG GGCGTGGAAC GCGACCTGAA CGGACGCGAC TACACGGTAA
2951  CGGGCGGCTT TACCGGCGCG ACTGCAGCAA CCGGCAAGAC GGGGGCACGC
3001  AATATGCCGC ACACCCGTCT GGTTGCCGGC CTGGGCGCGG ATGTCGAATT
3051  CGGCAACGGC TGGAACGGCT TGGCACGTTA CAGCTACGCC GGTTCCAAAC
3101  AGTACGGCAA CCACAGCGGA CGAGTCGGCG TAGGCTACCG GTTCCTCGAC
3151  GGTGGCGGAG GCACTGGATC CTCAGATTTG GCAAACGATT CTTTTATCCG
3201  GCAGGTTCTC GACCGTCAGC ATTTCGAACC CGACGGGAAA TACCACCTAT
3251  TCGGCAGCAG GGGGGAACTT GCCGAGCGCA GCGGCCATAT CGGATTGGGA
3301  AAAATACAAA GCCATCAGTT GGGCAACCTG ATGATTCAAC AGGCGGCCAT
3351  TAAAGGAAAT ATCGGCTACA TTGTCCGCTT TTCCGATCAC GGGCACGAAG
3401  TCCATTCCCC CTTCGACAAC CATGCCTCAC ATTCCGATTC TGATGAAGCC
3451  GGTAGTCCCG TTGACGGATT TAGCCTTTAC CGCATCCATT GGGACGGATA
3501  CGAACACCAT CCCGCCGACG GCTATGACGG GCCACAGGGC GGCGGCTATC
3551  CCGCTCCCAA AGGCGCGAGG GATATATACA GCTACGACAT AAAAGGCGTT
3601  GCCCAAAATA TCCGCCTCAA CCTGACCGAC AACCGCAGCA CCGGACAACG
3651  GCTTGCCGAC CGTTTCCACA ATGCCGGTAG TATGCTGACG CAAGGAGTAG
3701  GCGACGGATT CAAACGCGCC ACCCGATACA GCCCCGAGCT GGACAGATCG
3751  GGCAATGCCG CCGAAGCCTT CAACGGCACT GCAGATATCG TTAAAAACAT
3801  CATCGGCGCG GCAGGAGAAA TTGTCGGCGC AGGCGATGCC GTGCAGGGCA
3851  TAAGCGAAGG CTCAAACATT GCTGTCATGC ACGGCTTGGG TCTGCTTTCC
3901  ACCGAAAACA AGATGGCGCG CATCAACGAT TTGGCAGATA TGGCGCAACT
3951  CAAAGACTAT GCCGCAGCAG CCATCCGCGA TTGGGCAGTC CAAAACCCCA
4001  ATGCCGCACA AGGCATAGAA GCCGTCAGCA ATATCTTTAT GGCAGCCATC
4051  CCCATCAAAG GGATTGGAGC TGTTCGGGGA AAATACGGCT TGGGCGGCAT
4101  CACGGCACAT CCTATCAAGC GGTCGCAGAT GGGCGCGATC GCATTGCCGA
4151  AAGGGAAATC CGCCGTCAGC GACAATTTTG CCGATGCGGC ATACGCCAAA
4201  TACCCGTCCC CTTACCATTC CCGAAATATC CGTTCAAACT TGGAGCAGCG
4251  TTACGGCAAA GAAAACATCA CCTCCTCAAC CGTGCCGCCG TCAAACGGCA
4301  AAAATGTCAA ACTGGCAGAC CAACGCCACC CGAAGACAGG CGTACCGTTT
4351  GACGGTAAAG GGTTTCCGAA TTTTGAGAAG CACGTGAAAT ATGATACGCT
4401  CGAGCACCAC CACCACCACC ACTGA
```

```
   1  MTSAPDFNAG GTGIGSNSRA TTAKSAAVSY AGIKNEMCKD RSMLCAGRDD
  51  VAVTDRDAKI NAPPPNLHTG DFPNPNDAYK NLINLKPAIE AGYTGRGVEV
 101  GIVDTGESVG SISFPELYGR KEHGYNENYK NYTAYMRKEA PEDGGGKDIE
 151  ASFDDEAVIE TEAKPTDIRH VKEIGHIDLV SHIIGGRSVD GRPAGGIAPD
 201  ATLHIMNTND ETKNEMMVAA IRNAWVKLGE RGVRIVNNSF GTTSRAGTAD
 251  LFQIANSEEQ YRQALLDYSG GDKTDEGIRL MQQSDYGNLS YHIRNKNMLF
 301  IFSTGNDAQA QPNTYALLPF YEKDAQKGII TVAGVDRSGE KFKREMYGEP
 351  GTEPLEYGSN HCGITAMWCL SAPYEASVRF TRTNPIQIAG TSFSAPIVTG
 401  TAALLLQKYP WMSNDNLRTT LLTTAQDIGA VGVDSKFGWG LLDAGKAMNG
 451  PASFPFGDFT ADTKGTSDIA YSFRNDISGT GGLIKKGGSQ LQLHGNNTYT
 501  GKTIIEGGSL VLYGNNKSDM RVETKGALIY NGAASGGSLN SDGIVYLADT
 551  DQSGANETVH IKGSLQLDGK GTLYTRLGKL LKVDGTAIIG GKLYMSARGK
 601  GAGYLNSTGR RVPFLSAAKI GQDYSFFTNI ETDGGLLASL DSVEKTAGSE
 651  GDTLSYYVRR GNAARTASAA AHSAPAGLKH AVEQGGSNLE NLMVELDASE
 701  SSATPETVET AAADRTDMPG IRPYGATFRA AAAVQHANAA DGVRIFNSLA
 751  ATVYADSTAA HADMQGRRLK AVSDGLDHNG TGLRVIAQTQ QDGGTWEQGG
 801  VEGKMRGSTQ TVGIAAKTGE NTTAAATLGM GRSTWSENSA NAKTDSISLF
 851  AGIRHDAGDI GYLKGLFSYG RYKNSISRST GADEHAEGSV NGTLMQLGAL
 901  GGVNVPFAAT GDLTVEGGLR YDLLKQDAFA EKGSALGWSG NSLTEGTLVG
 951  LAGLKLSQPL SDKAVLFATA GVERDLNGRD YTVTGGFTGA TAATGKTGAR
1001  NMPHTRLVAG LGADVEFGNG WNGLARYSYA GSKQYGNHSG RVGVGYRFLD
1051  GGGGTGSSDL ANDSFIRQVL DRQHFEPDGK YHLFGSRGEL AERSGHIGLG
1101  KIQSHQLGNL MIQQAAIKGN IGYIVRFSDH GHEVHSPFDN HASHSDSDEA
1151  GSPVDGFSLY RIHWDGYEHH PADGYDGPQG GGYPAPKGAR DIYSYDIKGV
1201  AQNIRLNLTD NRSTGQRLAD RFHNAGSMLT QGVGDGFKRA TRYSPELDRS
1251  GNAAEAFNGT ADIVKNIIGA AGEIVGAGDA VQGISEGSNI AVMHGLGLLS
1301  TENKMARIND LADMAQLKDY AAAAIRDWAV QNPNAAQGIE AVSNIFMAAI
1351  PIKGIGAVRG KYGLGGITAH PIKRSQMGAI ALPKGKSAVS DNFADAAYAK
1401  YPSPYHSRNI RSNLEQRYGK ENITSSTVPP SNGKNVKLAD QRHPKTGVPF
1451  DGKGFPNFEK HVKYDTLEHH HHHH*
```

ΔG983-741

```
   1  ATGACTTCTG CGCCCGACTT CAATGCAGGC GGTACCGGTA TCGGCAGCAA
  51  CAGCAGAGCA ACAACAGCGA AATCAGCAGC AGTATCTTAC GCCGGTATCA
 101  AGAACGAAAT GTGCAAAGAC AGAAGCATGC TCTGTGCCGG TCGGGATGAC
 151  GTTGCGGTTA CAGACAGGGA TGCCAAAATC AATGCCCCCC CCCCGAATCT
 201  GCATACCGGA GACTTTCCAA ACCCAAATGA CGCATACAAG AATTTGATCA
 251  ACCTCAAACC TGCAATTGAA GCAGGCTATA CAGGACGCGG GGTAGAGGTA
 301  GGTATCGTCG ACACAGGCGA ATCCGTCGGC AGCATATCCT TTCCCGAACT
 351  GTATGGCAGA AAAGAACACG GCTATAACGA AAATTACAAA AACTATACGG
 401  CGTATATGCG GAAGGAAGCG CCTGAAGACG GAGGCGGTAA AGACATTGAA
 451  GCTTCTTTCG ACGATGAGGC CGTTATAGAG ACTGAAGCAA AGCCGACGGA
 501  TATCCGCCAC GTAAAAGAAA TCGGACACAT CGATTTGGTC TCCCATATTA
 551  TTGGCGGGCG TTCCGTGGAC GGCAGACCTG CAGGCGGTAT TGCGCCCGAT
 601  GCGACGCTAC ACATAATGAA TACGAATGAT GAAACCAAGA ACGAAATGAT
 651  GGTTGCAGCC ATCCGCAATG CATGGGTCAA GCTGGGCGAA CGTGGCGTGC
 701  GCATCGTCAA TAACAGTTTT GGAACAACAT CGAGGGCAGG CACTGCCGAC
 751  CTTTTCCAAA TAGCCAATTC GGAGGAGCAG TACCGCCAAG CGTTGCTCGA
 801  CTATTCCGGC GGTGATAAAA CAGACGAGGG TATCCGCCTG ATGCAACAGA
 851  GCGATTACGG CAACCTGTCC TACCACATCC GTAATAAAAA CATGCTTTTC
 901  ATCTTTTCGA CAGGCAATGA CGCACAAGCT CAGCCCAACA CATATGCCCT
 951  ATTGCCATTT TATGAAAAAG ACGCTCAAAA AGGCATTATC ACAGTCGCAG
1001  GCGTAGACCG CAGTGGAGAA AAGTTCAAAC GGGAAATGTA TGGAGAACCG
1051  GGTACAGAAC CGCTTGAGTA TGGCTCCAAC CATTGCGGAA TTACTGCCAT
1101  GTGGTGCCTG TCGGCACCCT ATGAAGCAAG CGTCCGTTTC ACCCGTACAA
1151  ACCCGATTCA AATTGCCGGA ACATCCTTTT CCGCACCCAT CGTAACCGGC
1201  ACGGCGGCTC TGCTGCTGCA GAAATACCCG TGGATGAGCA ACGACAACCT
1251  GCGTACCACG TTGCTGACGA CGGCTCAGGA CATCGGTGCA GTCGGCGTGG
1301  ACAGCAAGTT CGGCTGGGGA CTGCTGGATG CGGGTAAGGC CATGAACGGA
1351  CCCGCGTCCT TTCCGTTCGG CGACTTTACC GCCGATACGA AAGGTACATC
1401  CGATATTGCC TACTCCTTCC GTAACGACAT TTCAGGCACG GGCGGCCTGA
1451  TCAAAAAAGG CGGCAGCCAA CTGCAACTGC ACGGCAACAA CACCTATACG
1501  GGCAAAACCA TTATCGAAGG CGGTTCGCTG GTGTTGTACG GCAACAACAA
1551  ATCGGATATG CGCGTCGAAA CCAAAGGTGC GCTGATTTAT AACGGGGCGG
1601  CATCCGGCGG CAGCCTGAAC AGCGACGGCA TTGTCTATCT GGCAGATACC
1651  GACCAATCCG GCGCAAACGA AACCGTACAC ATCAAAGGCA GTCTGCAGCT
```

```
1701  GGACGGCAAA GGTACGCTGT ACACACGTTT GGGCAAACTG CTGAAAGTGG
1751  ACGGTACGGC GATTATCGGC GGCAAGCTGT ACATGTCGGC ACGCGGCAAG
1801  GGGGCAGGCT ATCTCAACAG TACCGGACGA CGTGTTCCCT TCCTGAGTGC
1851  CGCCAAAATC GGGCAGGATT ATTCTTTCTT CACAAACATC GAAACCGACG
1901  GCGGCCTGCT GGCTTCCCTC GACAGCGTCG AAAAAACAGC GGGCAGTGAA
1951  GGCGACACGC TGTCCTATTA TGTCCGTCGC GGCAATGCGG CACGGACTGC
2001  TTCGGCAGCG GCACATTCCG CGCCCGCCGG TCTGAAACAC GCCGTAGAAC
2051  AGGGCGGCAG CAATCTGGAA AACCTGATGG TCGAACTGGA TGCCTCCGAA
2101  TCATCCGCAA CACCCGAGAC GGTTGAAACT GCGGCAGCCG ACCGCACAGA
2151  TATGCCGGGC ATCCGCCCCT ACGGCGCAAC TTTCCGCGCA GCGGCAGCCG
2201  TACAGCATGC GAATGCCGCC GACGGTGTAC GCATCTTCAA CAGTCTCGCC
2251  GCTACCGTCT ATGCCGACAG TACCGCCGCC CATGCCGATA TGCAGGGACG
2301  CCGCCTGAAA GCCGTATCGG ACGGGTTGGA CCACAACGGC ACGGGTCTGC
2351  GCGTCATCGC GCAAACCCAA CAGGACGGTG GAACGTGGGA ACAGGGCGGT
2401  GTTGAAGGCA AAATGCGCGG CAGTACCCAA ACCGTCGGCA TTGCCGCGAA
2451  AACCGGCGAA AATACGACAG CAGCCGCCAC ACTGGGCATG GGACGCAGCA
2501  CATGGAGCGA AAACAGTGCA AATGCAAAAA CCGACAGCAT TAGTCTGTTT
2551  GCAGGCATAC GGCACGATGC GGGCGATATC GGCTATCTCA AAGGCCTGTT
2601  CTCCTACGGA CGCTACAAAA ACAGCATCAG CCGCAGCACC GGTGCGGACG
2651  AACATGCGGA AGGCAGCGTC AACGGCACGC TGATGCAGCT GGGCGCACTG
2701  GGCGGTGTCA ACGTTCCGTT TGCCGCAACG GGAGATTTGA CGGTCGAAGG
2751  CGGTCTGCGC TACGACCTGC TCAAACAGGA TGCATTCGCC GAAAAAGGCA
2801  GTGCTTTGGG CTGGAGCGGC AACAGCCTCA CTGAAGGCAC GCTGGTCGGA
2851  CTCGCGGGTC TGAAGCTGTC GCAACCCTTG AGCGATAAAG CCGTCCTGTT
2901  TGCAACGGCG GGCGTGGAAC GCGACCTGAA CGGACGCGAC TACACGGTAA
2951  CGGGCGGCTT TACCGGCGCG ACTGCAGCAA CCGGCAAGAC GGGGGCACGC
3001  AATATGCCGC ACACCCGTCT GGTTGCCGGC CTGGGCGCGG ATGTCGAATT
3051  CGGCAACGGC TGGAACGGCT TGGCACGTTA CAGCTACGCC GGTTCCAAAC
3101  AGTACGGCAA CCACAGCGGA CGAGTCGGCG TAGGCTACCG GTTCCTCGAG
3151  GGATCCGGAG GGGGTGGTGT CGCCGCCGAC ATCGGTGCGG GGCTTGCCGA
3201  TGCACTAACC GCACCGCTCG ACCATAAAGA CAAAGGTTTG CAGTCTTTGA
3251  CGCTGGATCA GTCCGTCAGG AAAAACGAGA AACTGAAGCT GGCGGCACAA
3301  GGTGCGGAAA AAACTTATGG AAACGGTGAC AGCCTCAATA CGGGCAAATT
3351  GAAGAACGAC AAGGTCAGCC GTTTCGACTT TATCCGCCAA ATCGAAGTGG
3401  ACGGGCAGCT CATTACCTTG GAGAGTGGAG AGTTCCAAGT ATACAAACAA
3451  AGCCATTCCG CCTTAACCGC CTTTCAGACC GAGCAAATAC AAGATTCGGA
3501  GCATTCCGGG AAGATGGTTG CGAAACGCCA GTTCAGAATC GGCGACATAG
3551  CGGGCGAACA TACATCTTTT GACAAGCTTC CCGAAGGCGG CAGGGCGACA
3601  TATCGCGGGA CGGCGTTCGG TTCAGACGAT GCCGGCGGAA AACTGACCTA
3651  CACCATAGAT TTCGCCGCCA AGCAGGGAAA CGGCAAAATC GAACATTTGA
3701  AATCGCCAGA ACTCAATGTC GACCTGGCCG CCGCCGATAT CAAGCCGGAT
3751  GGAAAACGCC ATGCCGTCAT CAGCGGTTCC GTCCTTTACA ACCAAGCCGA
3801  GAAAGGCAGT TACTCCCTCG GTATCTTTGG CGGAAAAGCC CAGGAAGTTG
3851  CCGGCAGCGC GGAAGTGAAA ACCGTAAACG GCATACGCCA TATCGGCCTT
3901  GCCGCCAAGC AACTCGAGCA CCACCACCAC CACCACTGA

   1  MTSAPDFNAG GTGIGSNSRA TTAKSAAVSY AGIKNEMCKD RSMLCAGRDD
  51  VAVTDRDAKI NAPPPNLHTG DFPNPNDAYK NLINLKPAIE AGYTGRGVEV
 101  GIVDTGESVG SISFPELYGR KEHGYNENYK NYTAYMRKEA PEDGGGKDIE
 151  ASFDDEAVIE TEAKPTDIRH VKEIGHIDLV SHIIGGRSVD GRPAGGIAPD
 201  ATLHIMNTND ETKNEMMVAA IRNAWVKLGE RGVRIVNNSF GTTSRAGTAD
 251  LFQIANSEEQ YRQALLDYSG GDKTDEGIRL MQQSDYGNLS YHIRNKNMLF
 301  IFSTGNDAQA QPNTYALLPF YEKDAQKGII TVAGVDRSGE KFKREMYGEP
 351  GTEPLEYGSN HCGITAMWCL SAPYEASVRF TRTNPIQIAG TSFSAPIVTG
 401  TAALLLQKYP WMSNDNLRTT LLTTAQDIGA VGVDSKFGWG LLDAGKAMNG
 451  PASFPFGDFT ADTKGTSDIA YSFRNDISGT GGLIKKGGSQ LQLHGNNTYT
 501  GKTIIEGGSL VLYGNNKSDM RVETKGALIY NGAASGGSLN SDGIVYLADT
 551  DQSGANETVH IKGSLQLDGK GTLYTRLGKL LKVDGTAIIG GKLYMSARGK
 601  GAGYLNSTGR RVPFLSAAKI GQDYSFFTNI ETDGGLLASL DSVEKTAGSE
 651  GDTLSYYVRR GNAARTASAA AHSAPAGLKH AVEQGGSNLE NLMVELDASE
 701  SSATPETVET AAADRTDMPG IRPYGATFRA AAAVQHANAA DGVRIFNSLA
 751  ATVYADSTAA HADMQGRRLK AVSDGLDHNG TGLRVIAQTQ QDGGTWEQGG
 801  VEGKMRGSTQ TVGIAAKTGE NTTAAATLGM GRSTWSENSA NAKTDSISLF
 851  AGIRHDAGDI GYLKGLFSYG RYKNSISRST GADEHAEGSV NGTLMQLGAL
 901  GGVNVPFAAT GDLTVEGGLR YDLLKQDAFA EKGSALGWSG NSLTEGTLVG
 951  LAGLKLSQPL SDKAVLFATA GVERDLNGRD YTVTGGFTGA TAATGKTGAR
1001  NMPHTRLVAG LGADVEFGNG WNGLARYSYA GSKQYGNHSG RVGVGYRFLE
```

```
1051   GSGGGGVAAD IGAGLADALT APLDHKDKGL QSLTLDQSVR KNEKLKLAAQ
1101   GAEKTYGNGD SLNTGKLKND KVSRFDFIRQ IEVDGQLITL ESGEFQVYKQ
1151   SHSALTAFQT EQIQDSEHSG KMVAKRQFRI GDIAGEHTSF DKLPEGGRAT
1201   YRGTAFGSDD AGGKLTYTID FAAKQGNGKI EHLKSPELNV DLAAADIKPD
1251   GKRHAVISGS VLYNQAEKGS YSLGIFGGKA QEVAGSAEVK TVNGIRHIGL
1301   AAKQLEHHHH HH*
```

ΔG983-961

```
   1 ATGACTTCTG CGCCCGACTT CAATGCAGGC GGTACCGGTA TCGGCAGCAA
  51 CAGCAGAGCA ACAACAGCGA AATCAGCAGC AGTATCTTAC GCCGGTATCA
 101 AGAACGAAAT GTGCAAAGAC AGAAGCATGC TCTGTGCCGG TCGGGATGAC
 151 GTTGCGGTTA CAGACAGGGA TGCCAAAATC AATGCCCCCC CCCCGAATCT
 201 GCATACCGGA GACTTTCCAA ACCCAAATGA CGCATACAAG AATTTGATCA
 251 ACCTCAAACC TGCAATTGAA GCAGGCTATA CAGGACGCGG GGTAGAGGTA
 301 GGTATCGTCG ACACAGGCGA ATCCGTCGGC AGCATATCCT TTCCCGAACT
 351 GTATGGCAGA AAAGAACACG GCTATAACGA AAATTACAAA AACTATACGG
 401 CGTATATGCG GAAGGAAGCG CCTGAAGACG GAGGCGGTAA AGACATTGAA
 451 GCTTCTTTCG ACGATGAGGC CGTTATAGAG ACTGAAGCAA AGCCGACGGA
 501 TATCCGCCAC GTAAAAGAAA TCGGACACAT CGATTTGGTC TCCCATATTA
 551 TTGGCGGGCG TTCCGTGGAC GGCAGACCTG CAGGCGGTAT TGCGCCCGAT
 601 GCGACGCTAC ACATAATGAA TACGAATGAT GAAACCAAGA ACGAAATGAT
 651 GGTTGCAGCC ATCCGCAATG CATGGGTCAA GCTGGGCGAA CGTGGCGTGC
 701 GCATCGTCAA TAACAGTTTT GGAACAACAT CGAGGGCAGG CACTGCCGAC
 751 CTTTTCCAAA TAGCCAATTC GGAGGAGCAG TACCGCCAAG CGTTGCTCGA
 801 CTATTCCGGC GGTGATAAAA CAGACGAGGG TATCCGCCTG ATGCAACAGA
 851 GCGATTACGG CAACCTGTCC TACCACATCC GTAATAAAAA CATGCTTTTC
 901 ATCTTTTCGA CAGGCAATGA CGCACAAGCT CAGCCCAACA CATATGCCCT
 951 ATTGCCATTT TATGAAAAAG ACGCTCAAAA AGGCATTATC ACAGTCGCAG
1001 GCGTAGACCG CAGTGGAGAA AAGTTCAAAC GGGAAATGTA TGGAGAACCG
1051 GGTACAGAAC CGCTTGAGTA TGGCTCCAAC CATTGCGGAA TTACTGCCAT
1101 GTGGTGCCTG TCGGCACCCT ATGAAGCAAG CGTCCGTTTC ACCCGTACAA
1151 ACCCGATTCA AATTGCCGGA ACATCCTTTT CCGCACCCAT CGTAACCGGC
1201 ACGGCGGCTC TGCTGCTGCA GAAATACCCG TGGATGAGCA ACGACAACCT
1251 GCGTACCACG TTGCTGACGA CGGCTCAGGA CATCGGTGCA GTCGGCGTGG
1301 ACAGCAAGTT CGGCTGGGGA CTGCTGGATG CGGGTAAGGC CATGAACGGA
1351 CCCGCGTCCT TTCCGTTCGG CGACTTTACC GCCGATACGA AAGGTACATC
1401 CGATATTGCC TACTCCTTCC GTAACGACAT TTCAGGCACG GGCGGCCTGA
1451 TCAAAAAAGG CGGCAGCCAA CTGCAACTGC ACGGCAACAA CACCTATACG
1501 GGCAAAACCA TTATCGAAGG CGGTTCGCTG GTGTTGTACG GCAACAACAA
1551 ATCGGATATG CGCGTCGAAA CCAAAGGTGC GCTGATTTAT AACGGGGCGG
1601 CATCCGGCGG CAGCCTGAAC AGCGACGGCA TTGTCTATCT GGCAGATACC
1651 GACCAATCCG GCGCAAACGA AACCGTACAC ATCAAAGGCA GTCTGCAGCT
1701 GGACGGCAAA GGTACGCTGT ACACACGTTT GGGCAAACTG CTGAAAGTGG
1751 ACGGTACGGC GATTATCGGC GGCAAGCTGT ACATGTCGGC ACGCGGCAAG
1801 GGGGCAGGCT ATCTCAACAG TACCGGACGA CGTGTTCCCT TCCTGAGTGC
1851 CGCCAAAATC GGGCAGGATT ATTCTTTCTT CACAAACATC GAAACCGACG
1901 GCGGCCTGCT GGCTTCCCTC GACAGCGTCG AAAAAACAGC GGGCAGTGAA
1951 GGCGACACGC TGTCCTATTA TGTCCGTCGC GGCAATGCGG CACGGACTGC
2001 TTCGGCAGCG GCACATTCCG CGCCCGCCGG TCTGAAACAC GCCGTAGAAC
2051 AGGGCGGCAG CAATCTGGAA AACCTGATGG TCGAACTGGA TGCCTCCGAA
2101 TCATCCGCAA CACCCGAGAC GGTTGAAACT GCGGCAGCCG ACCGCACAGA
2151 TATGCCGGGC ATCCGCCCCT ACGGCGCAAC TTTCCGCGCA GCGGCAGCCG
2201 TACAGCATGC GAATGCCGCC GACGGTGTAC GCATCTTCAA CAGTCTCGCC
2251 GCTACCGTCT ATGCCGACAG TACCGCCGCC CATGCCGATA TGCAGGGACG
2301 CCGCCTGAAA GCCGTATCGG ACGGGTTGGA CCACAACGGC ACGGGTCTGC
2351 GCGTCATCGC GCAAACCCAA CAGGACGGTG GAACGTGGGA ACAGGGCGGT
2401 GTTGAAGGCA AAATGCGCGG CAGTACCCAA ACCGTCGGCA TTGCCGCGAA
2451 AACCGGCGAA AATACGACAG CAGCCGCCAC ACTGGGCATG GGACGCAGCA
2501 CATGGAGCGA AAACAGTGCA AATGCAAAAA CCGACAGCAT TAGTCTGTTT
2551 GCAGGCATAC GGCACGATGC GGGCGATATC GGCTATCTCA AAGGCCTGTT
2601 CTCCTACGGA CGCTACAAAA ACAGCATCAG CCGCAGCACC GGTGCGGACG
2651 AACATGCGGA AGGCAGCGTC AACGGCACGC TGATGCAGCT GGGCGCACTG
2701 GGCGGTGTCA ACGTTCCGTT TGCCGCAACG GGAGATTTGA CGGTCGAAGG
2751 CGGTCTGCGC TACGACCTGC TCAAACAGGA TGCATTCGCC GAAAAAGGCA
2801 GTGCTTTGGG CTGGAGCGGC AACAGCCTCA CTGAAGGCAC GCTGGTCGGA
2851 CTCGCGGGTC TGAAGCTGTC GCAACCCTTG AGCGATAAAG CCGTCCTGTT
```

```
2901   TGCAACGGCG GGCGTGGAAC GCGACCTGAA CGGACGCGAC TACACGGTAA
2951   CGGGCGGCTT TACCGGCGCG ACTGCAGCAA CCGGCAAGAC GGGGGCACGC
3001   AATATGCCGC ACACCCGTCT GGTTGCCGGC CTGGGCGCGG ATGTCGAATT
3051   CGGCAACGGC TGGAACGGCT TGGCACGTTA CAGCTACGCC GGTTCCAAAC
3101   AGTACGGCAA CCACAGCGGA CGAGTCGGCG TAGGCTACCG GTTCCTCGAG
3151   GGTGGCGGAG GCACTGGATC CGCCACAAAC GACGACGATG TTAAAAAAGC
3201   TGCCACTGTG GCCATTGCTG CTGCCTACAA CAATGGCCAA GAAATCAACG
3251   GTTTCAAAGC TGGAGAGACC ATCTACGACA TTGATGAAGA CGGCACAATT
3301   ACCAAAAAAG ACGCAACTGC AGCCGATGTT GAAGCCGACG ACTTTAAAGG
3351   TCTGGGTCTG AAAAAAGTCG TGACTAACCT GACCAAAACC GTCAATGAAA
3401   ACAAACAAAA CGTCGATGCC AAAGTAAAAG CTGCAGAATC TGAAATAGAA
3451   AAGTTAACAA CCAAGTTAGC AGACACTGAT GCCGCTTTAG CAGATACTGA
3501   TGCCGCTCTG GATGCAACCA CCAACGCCTT GAATAAATTG GGAGAAAATA
3551   TAACGACATT TGCTGAAGAG ACTAAGACAA ATATCGTAAA AATTGATGAA
3601   AAATTAGAAG CCGTGGCTGA TACCGTCGAC AAGCATGCCG AAGCATTCAA
3651   CGATATCGCC GATTCATTGG ATGAAACCAA CACTAAGGCA GACGAAGCCG
3701   TCAAAACCGC CAATGAAGCC AAACAGACGG CCGAAGAAAC CAAACAAAAC
3751   GTCGATGCCA AAGTAAAAGC TGCAGAAACT GCAGCAGGCA AAGCCGAAGC
3801   TGCCGCTGGC ACAGCTAATA CTGCAGCCGA CAAGGCCGAA GCTGTCGCTG
3851   CAAAAGTTAC CGACATCAAA GCTGATATCG CTACGAACAA AGATAATATT
3901   GCTAAAAAAG CAAACAGTGC CGACGTGTAC ACCAGAGAAG AGTCTGACAG
3951   CAAATTTGTC AGAATTGATG GTCTGAACGC TACTACCGAA AAATTGGACA
4001   CACGCTTGGC TTCTGCTGAA AAATCCATTG CCGATCACGA TACTCGCCTG
4051   AACGGTTTGG ATAAAACAGT GTCAGACCTG CGCAAAGAAA CCCGCCAAGG
4101   CCTTGCAGAA CAAGCCGCGC TCTCCGGTCT GTTCCAACCT TACAACGTGG
4151   GTCGGTTCAA TGTAACGGCT GCAGTCGGCG GCTACAAATC CGAATCGGCA
4201   GTCGCCATCG GTACCGGCTT CCGCTTTACC GAAAACTTTG CCGCCAAAGC
4251   AGGCGTGGCA GTCGGCACTT CGTCCGGTTC TTCCGCAGCC TACCATGTCG
4301   GCGTCAATTA CGAGTGGCTC GAGCACCACC ACCACCACCA CTGA

1      MTSAPDFNAG GTGIGSNSRA TTAKSAAVSY AGIKNEMCKD RSMLCAGRDD
51     VAVTDRDAKI NAPPPNLHTG DFPNPNDAYK NLINLKPAIE AGYTGRGVEV
101    GIVDTGESVG SISFPELYGR KEHGYNENYK NYTAYMRKEA PEDGGGKDIE
151    ASFDDEAVIE TEAKPTDIRH VKEIGHIDLV SHIIGGRSVD GRPAGGIAPD
201    ATLHIMNTND ETKNEMMVAA IRNAWVKLGE RGVRIVNNSF GTTSRAGTAD
251    LFQIANSEEQ YRQALLDYSG GDKTDEGIRL MQQSDYGNLS YHIRNKNMLF
301    IFSTGNDAQA QPNTYALLPF YEKDAQKGII TVAGVDRSGE KFKREMYGEP
351    GTEPLEYGSN HCGITAMWCL SAPYEASVRF TRTNPIQIAG TSFSAPIVTG
401    TAALLLQKYP WMSNDNLRTT LLTTAQDIGA VGVDSKFGWG LLDAGKAMNG
451    PASFPFGDFT ADTKGTSDIA YSFRNDISGT GGLIKKGGSQ LQLHGNNTYT
501    GKTIIEGGSL VLYGNNKSDM RVETKGALIY NGAASGGSLN SDGIVYLADT
551    DQSGANETVH IKGSLQLDGK GTLYTRLGKL LKVDGTAIIG GKLYMSARGK
601    GAGYLNSTGR RVPFLSAAKI GQDYSFFTNI ETDGGLLASL DSVEKTAGSE
651    GDTLSYYVRR GNAARTASAA AHSAPAGLKH AVEQGGSNLE NLMVELDASE
701    SSATPETVET AAADRTDMPG IRPYGATFRA AAAVQHANAA DGVRIFNSLA
751    ATVYADSTAA HADMQGRRLK AVSDGLDHNG TGLRVIAQTQ QDGGTWEQGG
801    VEGKMRGSTQ TVGIAAKTGE NTTAAATLGM GRSTWSENSA NAKTDSISLF
851    AGIRHDAGDI GYLKGLFSYG RYKNSISRST GADEHAEGSV NGTLMQLGAL
901    GGVNVPFAAT GDLTVEGGLR YDLLKQDAFA EKGSALGWSG NSLTEGTLVG
951    LAGLKLSQPL SDKAVLFATA GVERDLNGRD YTVTGGFTGA TAATGKTGAR
1001   NMPHTRLVAG LGADVEFGNG WNGLARYSYA GSKQYGNHSG RVGVGYRFLE
1051   GGGGTGSATN DDDVKKAATV AIAAAYNNGQ EINGFKAGET IYDIDEDGTI
1101   TKKDATAADV EADDFKGLGL KKVVTNLTKT VNENKQNVDA KVKAAESEIE
1151   KLTTKLADTD AALADTDAAL DATTNALNKL GENITTFAEE TKTNIVKIDE
1201   KLEAVADTVD KHAEAFNDIA DSLDETNTKA DEAVKTANEA KQTAEETKQN
1251   VDAKVKAAET AAGKAEAAAG TANTAADKAE AVAAKVTDIK ADIATNKDNI
1301   AKKANSADVY TREESDSKFV RIDGLNATTE KLDTRLASAE KSIADHDTRL
1351   NGLDKTVSDL RKETRQGLAE QAALSGLFQP YNVGRFNVTA AVGGYKSESA
1401   VAIGTGFRFT ENFAAKAGVA VGTSSGSSAA YHVGVNYEWL EHHHHHH*

ΔG983-961c
1      ATGACTTCTG CGCCCGACTT CAATGCAGGC GGTACCGGTA TCGGCAGCAA
51     CAGCAGAGCA ACAACAGCGA AATCAGCAGC AGTATCTTAC GCCGGTATCA
101    AGAACGAAAT GTGCAAAGAC AGAAGCATGC TCTGTGCCGG TCGGGATGAC
151    GTTGCGGTTA CAGACAGGGA TGCCAAAATC AATGCCCCCC CCCCGAATCT
201    GCATACCGGA GACTTTCCAA ACCCAAATGA CGCATACAAG AATTTGATCA
```

```
 251  ACCTCAAACC TGCAATTGAA GCAGGCTATA CAGGACGCGG GGTAGAGGTA
 301  GGTATCGTCG ACACAGGCGA ATCCGTCGGC AGCATATCCT TTCCCGAACT
 351  GTATGGCAGA AAAGAACACG GCTATAACGA AAATTACAAA AACTATACGG
 401  CGTATATGCG GAAGGAAGCG CCTGAAGACG GAGGCGGTAA AGACATTGAA
 451  GCTTCTTTCG ACGATGAGGC CGTTATAGAG ACTGAAGCAA AGCCGACGGA
 501  TATCCGCCAC GTAAAAGAAA TCGGACACAT CGATTTGGTC TCCCATATTA
 551  TTGGCGGGCG TTCCGTGGAC GGCAGACCTG CAGGCGGTAT TGCGCCCGAT
 601  GCGACGCTAC ACATAATGAA TACGAATGAT GAAACCAAGA ACGAAATGAT
 651  GGTTGCAGCC ATCCGCAATG CATGGGTCAA GCTGGGCGAA CGTGGCGTGC
 701  GCATCGTCAA TAACAGTTTT GGAACAACAT CGAGGGCAGG CACTGCCGAC
 751  CTTTTCCAAA TAGCCAATTC GGAGGAGCAG TACCGCCAAG CGTTGCTCGA
 801  CTATTCCGGC GGTGATAAAA CAGACGAGGG TATCCGCCTG ATGCAACAGA
 851  GCGATTACGG CAACCTGTCC TACCACATCC GTAATAAAAA CATGCTTTTC
 901  ATCTTTTCGA CAGGCAATGA CGCACAAGCT CAGCCCAACA CATATGCCCT
 951  ATTGCCATTT TATGAAAAAG ACGCTCAAAA AGGCATTATC ACAGTCGCAG
1001  GCGTAGACCG CAGTGGAGAA AAGTTCAAAC GGGAAATGTA TGGAGAACCG
1051  GGTACAGAAC CGCTTGAGTA TGGCTCCAAC CATTGCGGAA TTACTGCCAT
1101  GTGGTGCCTG TCGGCACCCT ATGAAGCAAG CGTCCGTTTC ACCCGTACAA
1151  ACCCGATTCA AATTGCCGGA ACATCCTTTT CCGCACCCAT CGTAACCGGC
1201  ACGGCGGCTC TGCTGCTGCA GAAATACCCG TGGATGAGCA ACGACAACCT
1251  GCGTACCACG TTGCTGACGA CGGCTCAGGA CATCGGTGCA GTCGGCGTGG
1301  ACAGCAAGTT CGGCTGGGGA CTGCTGGATG CGGGTAAGGC CATGAACGGA
1351  CCCGCGTCCT TTCCGTTCGG CGACTTTACC GCCGATACGA AAGGTACATC
1401  CGATATTGCC TACTCCTTCC GTAACGACAT TTCAGGCACG GGCGGCCTGA
1451  TCAAAAAAGG CGGCAGCCAA CTGCAACTGC ACGGCAACAA CACCTATACG
1501  GGCAAAACCA TTATCGAAGG CGGTTCGCTG GTGTTGTACG GCAACAACAA
1551  ATCGGATATG CGCGTCGAAA CCAAAGGTGC GCTGATTTAT AACGGGGCGG
1601  CATCCGGCGG CAGCCTGAAC AGCGACGGCA TTGTCTATCT GGCAGATACC
1651  GACCAATCCG GCGCAAACGA AACCGTACAC ATCAAAGGCA GTCTGCAGCT
1701  GGACGGCAAA GGTACGCTGT ACACACGTTT GGGCAAACTG CTGAAAGTGG
1751  ACGGTACGGC GATTATCGGC GGCAAGCTGT ACATGTCGGC ACGCGGCAAG
1801  GGGGCAGGCT ATCTCAACAG TACCGGACGA CGTGTTCCCT TCCTGAGTGC
1851  CGCCAAAATC GGGCAGGATT ATTCTTTCTT CACAAACATC GAAACCGACG
1901  GCGGCCTGCT GGCTTCCCTC GACAGCGTCG AAAAAACAGC GGGCAGTGAA
1951  GGCGACACGC TGTCCTATTA TGTCCGTCGC GGCAATGCGG CACGGACTGC
2001  TTCGGCAGCG GCACATTCCG CGCCCGCCGG TCTGAAACAC GCCGTAGAAC
2051  AGGGCGGCAG CAATCTGGAA AACCTGATGG TCGAACTGGA TGCCTCCGAA
2101  TCATCCGCAA CACCCGAGAC GGTTGAAACT GCGGCAGCCG ACCGCACAGA
2151  TATGCCGGGC ATCCGCCCCT ACGGCGCAAC TTTCCGCGCA GCGGCAGCCG
2201  TACAGCATGC GAATGCCGCC GACGGTGTAC GCATCTTCAA CAGTCTCGCC
2251  GCTACCGTCT ATGCCGACAG TACCGCCGCC CATGCCGATA TGCAGGGACG
2301  CCGCCTGAAA GCCGTATCGG ACGGGTTGGA CCACAACGGC ACGGGTCTGC
2351  GCGTCATCGC GCAAACCCAA CAGGACGGTG GAACGTGGGA ACAGGGCGGT
2401  GTTGAAGGCA AAATGCGCGG CAGTACCCAA ACCGTCGGCA TTGCCGCGAA
2451  AACCGGCGAA AATACGACAG CAGCCGCCAC ACTGGGCATG GGACGCAGCA
2501  CATGGAGCGA AAACAGTGCA AATGCAAAAA CCGACAGCAT TAGTCTGTTT
2551  GCAGGCATAC GGCACGATGC GGGCGATATC GGCTATCTCA AAGGCCTGTT
2601  CTCCTACGGA CGCTACAAAA ACAGCATCAG CCGCAGCACC GGTGCGGACG
2651  AACATGCGGA AGGCAGCGTC AACGGCACGC TGATGCAGCT GGGCGCACTG
2701  GGCGGTGTCA ACGTTCCGTT TGCCGCAACG GGAGATTTGA CGGTCGAAGG
2751  CGGTCTGCGC TACGACCTGC TCAAACAGGA TGCATTCGCC GAAAAAGGCA
2801  GTGCTTTGGG CTGGAGCGGC AACAGCCTCA CTGAAGGCAC GCTGGTCGGA
2851  CTCGCGGGTC TGAAGCTGTC GCAACCCTTG AGCGATAAAG CCGTCCTGTT
2901  TGCAACGGCG GGCGTGGAAC GCGACCTGAA CGGACGCGAC TACACGGTAA
2951  CGGGCGGCTT TACCGGCGCG ACTGCAGCAA CCGGCAAGAC GGGGGCACGC
3001  AATATGCCGC ACACCCGTCT GGTTGCCGGC CTGGGCGCGG ATGTCGAATT
3051  CGGCAACGGC TGGAACGGCT TGGCACGTTA CAGCTACGCC GGTTCCAAAC
3101  AGTACGGCAA CCACAGCGGA CGAGTCGGCG TAGGCTACCG GTTCCTCGAG
3151  GGTGGCGGAG GCACTGGATC CGCCACAAAC GACGACGATG TTAAAAAAGC
3201  TGCCACTGTG GCCATTGCTG CTGCCTACAA CAATGGCCAA GAAATCAACG
3251  GTTTCAAAGC TGGAGAGACC ATCTACGACA TTGATGAAGA CGGCACAATT
3301  ACCAAAAAAG ACGCAACTGC AGCCGATGTT GAAGCCGACG ACTTTAAAGG
3351  TCTGGGTCTG AAAAAAGTCG TGACTAACCT GACCAAAACC GTCAATGAAA
3401  ACAAACAAAA CGTCGATGCC AAAGTAAAAG CTGCAGAATC TGAAATAGAA
3451  AAGTTAACAA CCAAGTTAGC AGACACTGAT GCCGCTTTAG CAGATACTGA
3501  TGCCGCTCTG GATGCAACCA CCAACGCCTT GAATAAATTG GGAGAAAATA
3551  TAACGACATT TGCTGAAGAG ACTAAGACAA ATATCGTAAA AATTGATGAA
```

```
3601   AAATTAGAAG CCGTGGCTGA TACCGTCGAC AAGCATGCCG AAGCATTCAA
3651   CGATATCGCC GATTCATTGG ATGAAACCAA CACTAAGGCA GACGAAGCCG
3701   TCAAAACCGC CAATGAAGCC AAACAGACGG CCGAAGAAAC CAAACAAAAC
3751   GTCGATGCCA AAGTAAAAGC TGCAGAAACT GCAGCAGGCA AAGCCGAAGC
3801   TGCCGCTGGC ACAGCTAATA CTGCAGCCGA CAAGGCCGAA GCTGTCGCTG
3851   CAAAAGTTAC CGACATCAAA GCTGATATCG CTACGAACAA AGATAATATT
3901   GCTAAAAAAG CAAACAGTGC CGACGTGTAC ACCAGAGAAG AGTCTGACAG
3951   CAAATTTGTC AGAATTGATG GTCTGAACGC TACTACCGAA AAATTGGACA
4001   CACGCTTGGC TTCTGCTGAA AAATCCATTG CCGATCACGA TACTCGCCTG
4051   AACGGTTTGG ATAAAACAGT GTCAGACCTG CGCAAAGAAA CCCGCCAAGG
4101   CCTTGCAGAA CAAGCCGCGC TCTCCGGTCT GTTCCAACCT TACAACGTGG
4151   GTCTCGAGCA CCACCACCAC CACCACTGA

   1   MTSAPDFNAG GTGIGSNSRA TTAKSAAVSY AGIKNEMCKD RSMLCAGRDD
  51   VAVTDRDAKI NAPPPNLHTG DFPNPNDAYK NLINLKPAIE AGYTGRGVEV
 101   GIVDTGESVG SISFPELYGR KEHGYNENYK NYTAYMRKEA PEDGGGKDIE
 151   ASFDDEAVIE TEAKPTDIRH VKEIGHIDLV SHIIGGRSVD GRPAGGIAPD
 201   ATLHIMNTND ETKNEMMVAA IRNAWVKLGE RGVRIVNNSF GTTSRAGTAD
 251   LFQIANSEEQ YRQALLDYSG GDKTDEGIRL MQQSDYGNLS YHIRNKNMLF
 301   IFSTGNDAQA QPNTYALLPF YEKDAQKGII TVAGVDRSGE KFKREMYGEP
 351   GTEPLEYGSN HCGITAMWCL SAPYEASVRF TRTNPIQIAG TSFSAPIVTG
 401   TAALLLQKYP WMSNDNLRTT LLTTAQDIGA VGVDSKFGWG LLDAGKAMNG
 451   PASFPFGDFT ADTKGTSDIA YSFRNDISGT GGLIKKGGSQ LQLHGNNTYT
 501   GKTIIEGGSL VLYGNNKSDM RVETKGALIY NGAASGGSLN SDGIVYLADT
 551   DQSGANETVH IKGSLQLDGK GTLYTRLGKL LKVDGTAIIG GKLYMSARGK
 601   GAGYLNSTGR RVPFLSAAKI GQDYSFFTNI ETDGGLLASL DSVEKTAGSE
 651   GDTLSYYVRR GNAARTASAA AHSAPAGLKH AVEQGGSNLE NLMVELDASE
 701   SSATPETVET AAADRTDMPG IRPYGATFRA AAAVQHANAA DGVRIFNSLA
 751   ATVYADSTAA HADMQGRRLK AVSDGLDHNG TGLRVIAQTQ QDGGTWEQGG
 801   VEGKMRGSTQ TVGIAAKTGE NTTAAATLGM GRSTWSENSA NAKTDSISLF
 851   AGIRHDAGDI GYLKGLFSYG RYKNSISRST GADEHAEGSV NGTLMQLGAL
 901   GGVNVPFAAT GDLTVEGGLR YDLLKQDAFA EKGSALGWSG NSLTEGTLVG
 951   LAGLKLSQPL SDKAVLFATA GVERDLNGRD YTVTGGFTGA TAATGKTGAR
1001   NMPHTRLVAG LGADVEFGNG WNGLARYSYA GSKQYGNHSG RVGVGYRFLE
1051   GGGGTGSATN DDDVKKAATV AIAAAYNNGQ EINGFKAGET IYDIDEDGTI
1101   TKKDATAADV EADDFKGLGL KKVVTNLTKT VNENKQNVDA KVKAAESEIE
1151   KLTTKLADTD AALADTDAAL DATTNALNKL GENITTFAEE TKTNIVKIDE
1201   KLEAVADTVD KHAEAFNDIA DSLDETNTKA DEAVKTANEA KQTAEETKQN
1251   VDAKVKAAET AAGKAEAAAG TANTAADKAE AVAAKVTDIK ADIATNKDNI
1301   AKKANSADVY TREESDSKFV RIDGLNATTE KLDTRLASAE KSIADHDTRL
1351   NGLDKTVSDL RKETRQGLAE QAALSGLFQP YNVGLEHHHH HH*
```

ΔG741 and hybrids

Bactericidal titres generated in response to ΔG741 (His-fusion) were measured against various strains, including the homologous 2996 strain:

	2996	MC58	NGH38	F6124	BZ133
ΔG741	512	131072	>2048	16384	>2048

As can be seen, the ΔG741-induced anti-bactericidal titre is particularly high against heterologous strain MC58.

ΔG741 was also fused directly in-frame upstream of proteins 961, 961c, 983 and ORF46.1:

ΔG741-961

```
   1   ATGGTCGCCG CCGACATCGG TGCGGGGCTT GCCGATGCAC TAACCGCACC
  51   GCTCGACCAT AAAGACAAAG GTTTGCAGTC TTTGACGCTG GATCAGTCCG
 101   TCAGGAAAAA CGAGAAACTG AAGCTGGCGG CACAAGGTGC GGAAAAAACT
 151   TATGGAAACG GTGACAGCCT CAATACGGGC AAATTGAAGA ACGACAAGGT
 201   CAGCCGTTTC GACTTTATCC GCCAAATCGA AGTGGACGGG CAGCTCATTA
```

```
 251  CCTTGGAGAG TGGAGAGTTC CAAGTATACA AACAAAGCCA TTCCGCCTTA
 301  ACCGCCTTTC AGACCGAGCA AATACAAGAT TCGGAGCATT CCGGGAAGAT
 351  GGTTGCGAAA CGCCAGTTCA GAATCGGCGA CATAGCGGGC GAACATACAT
 401  CTTTTGACAA GCTTCCCGAA GGCGGCAGGG CGACATATCG CGGGACGGCG
 451  TTCGGTTCAG ACGATGCCGG CGGAAAACTG ACCTACACCA TAGATTTCGC
 501  CGCCAAGCAG GGAAACGGCA AAATCGAACA TTTGAAATCG CCAGAACTCA
 551  ATGTCGACCT GGCCGCCGCC GATATCAAGC CGGATGGAAA ACGCCATGCC
 601  GTCATCAGCG GTTCCGTCCT TTACAACCAA GCCGAGAAAG GCAGTTACTC
 651  CCTCGGTATC TTTGGCGGAA AAGCCCAGGA AGTTGCCGGC AGCGCGGAAG
 701  TGAAAACCGT AAACGGCATA CGCCATATCG GCCTTGCCGC CAAGCAACTC
 751  GAGGGTGGCG GAGGCACTGG ATCCGCCACA AACGACGACG ATGTTAAAAA
 801  AGCTGCCACT GTGGCCATTG CTGCTGCCTA CAACAATGGC CAAGAAATCA
 851  ACGGTTTCAA AGCTGGAGAG ACCATCTACG ACATTGATGA AGACGGCACA
 901  ATTACCAAAA AAGACGCAAC TGCAGCCGAT GTTGAAGCCG ACGACTTTAA
 951  AGGTCTGGGT CTGAAAAAAG TCGTGACTAA CCTGACCAAA ACCGTCAATG
1001  AAAACAAACA AAACGTCGAT GCCAAAGTAA AAGCTGCAGA ATCTGAAATA
1051  GAAAAGTTAA CAACCAAGTT AGCAGACACT GATGCCGCTT TAGCAGATAC
1101  TGATGCCGCT CTGGATGCAA CCACCAACGC CTTGAATAAA TTGGGAGAAA
1151  ATATAACGAC ATTTGCTGAA GAGACTAAGA CAAATATCGT AAAAATTGAT
1201  GAAAAATTAG AAGCCGTGGC TGATACCGTC GACAAGCATG CCGAAGCATT
1251  CAACGATATC GCCGATTCAT TGGATGAAAC CAACACTAAG GCAGACGAAG
1301  CCGTCAAAAC CGCCAATGAA GCCAAACAGA CGGCCGAAGA AACCAAACAA
1351  AACGTCGATG CCAAAGTAAA AGCTGCAGAA ACTGCAGCAG GCAAAGCCGA
1401  AGCTGCCGCT GGCACAGCTA ATACTGCAGC CGACAAGGCC GAAGCTGTCG
1451  CTGCAAAAGT TACCGACATC AAAGCTGATA TCGCTACGAA CAAAGATAAT
1501  ATTGCTAAAA AAGCAAACAG TGCCGACGTG TACACCAGAG AAGAGTCTGA
1551  CAGCAAATTT GTCAGAATTG ATGGTCTGAA CGCTACTACC GAAAAATTGG
1601  ACACACGCTT GGCTTCTGCT GAAAAATCCA TTGCCGATCA CGATACTCGC
1651  CTGAACGGTT TGGATAAAAC AGTGTCAGAC CTGCGCAAAG AAACCCGCCA
1701  AGGCCTTGCA GAACAAGCCG CGCTCTCCGG TCTGTTCCAA CCTTACAACG
1751  TGGGTCGGTT CAATGTAACG GCTGCAGTCG GCGGCTACAA ATCCGAATCG
1801  GCAGTCGCCA TCGGTACCGG CTTCCGCTTT ACCGAAAACT TTGCCGCCAA
1851  AGCAGGCGTG GCAGTCGGCA CTTCGTCCGG TTCTTCCGCA GCCTACCATG
1901  TCGGCGTCAA TTACGAGTGG CTCGAGCACC ACCACCACCA CCACTGA

   1  MVAADIGAGL ADALTAPLDH KDKGLQSLTL DQSVRKNEKL KLAAQGAEKT
  51  YGNGDSLNTG KLKNDKVSRF DFIRQIEVDG QLITLESGEF QVYKQSHSAL
 101  TAFQTEQIQD SEHSGKMVAK RQFRIGDIAG EHTSFDKLPE GGRATYRGTA
 151  FGSDDAGGKL TYTIDFAAKQ GNGKIEHLKS PELNVDLAAA DIKPDGKRHA
 201  VISGSVLYNQ AEKGSYSLGI FGGKAQEVAG SAEVKTVNGI RHIGLAAKQL
 251  EGGGGTGSAT NDDDVKKAAT VAIAAAYNNG QEINGFKAGE TIYDIDEDGT
 301  ITKKDATAAD VEADDFKGLG LKKVVTNLTK TVNENKQNVD AKVKAAESEI
 351  EKLTTKLADT DAALADTDAA LDATTNALNK LGENITTFAE ETKTNIVKID
 401  EKLEAVADTV DKHAEAFNDI ADSLDETNTK ADEAVKTANE AKQTAEETKQ
 451  NVDAKVKAAE TAAGKAEAAA GTANTAADKA EAVAAKVTDI KADIATNKDN
 501  IAKKANSADV YTREESDSKF VRIDGLNATT EKLDTRLASA EKSIADHDTR
 551  LNGLDKTVSD LRKETRQGLA EQAALSGLFQ PYNVGRFNVT AAVGGYKSES
 601  AVAIGTGFRF TENFAAKAGV AVGTSSGSSA AYHVGVNYEW LEHHHHHH*
```

ΔG741–961c

```
   1  ATGGTCGCCG CCGACATCGG TGCGGGGCTT GCCGATGCAC TAACCGCACC
  51  GCTCGACCAT AAAGACAAAG GTTTGCAGTC TTTGACGCTG GATCAGTCCG
 101  TCAGGAAAAA CGAGAAACTG AAGCTGGCGG CACAAGGTGC GGAAAAAACT
 151  TATGGAAACG GTGACAGCCT CAATACGGGC AAATTGAAGA ACGACAAGGT
 201  CAGCCGTTTC GACTTTATCC GCCAAATCGA AGTGGACGGG CAGCTCATTA
 251  CCTTGGAGAG TGGAGAGTTC CAAGTATACA AACAAAGCCA TTCCGCCTTA
 301  ACCGCCTTTC AGACCGAGCA AATACAAGAT TCGGAGCATT CCGGGAAGAT
 351  GGTTGCGAAA CGCCAGTTCA GAATCGGCGA CATAGCGGGC GAACATACAT
 401  CTTTTGACAA GCTTCCCGAA GGCGGCAGGG CGACATATCG CGGGACGGCG
 451  TTCGGTTCAG ACGATGCCGG CGGAAAACTG ACCTACACCA TAGATTTCGC
 501  CGCCAAGCAG GGAAACGGCA AAATCGAACA TTTGAAATCG CCAGAACTCA
 551  ATGTCGACCT GGCCGCCGCC GATATCAAGC CGGATGGAAA ACGCCATGCC
 601  GTCATCAGCG GTTCCGTCCT TTACAACCAA GCCGAGAAAG GCAGTTACTC
 651  CCTCGGTATC TTTGGCGGAA AAGCCCAGGA AGTTGCCGGC AGCGCGGAAG
 701  TGAAAACCGT AAACGGCATA CGCCATATCG GCCTTGCCGC CAAGCAACTC
 751  GAGGGTGGCG GAGGCACTGG ATCCGCCACA AACGACGACG ATGTTAAAAA
```

```
 801  AGCTGCCACT GTGGCCATTG CTGCTGCCTA CAACAATGGC CAAGAAATCA
 851  ACGGTTTCAA AGCTGGAGAG ACCATCTACG ACATTGATGA AGACGGCACA
 901  ATTACCAAAA AAGACGCAAC TGCAGCCGAT GTTGAAGCCG ACGACTTTAA
 951  AGGTCTGGGT CTGAAAAAAG TCGTGACTAA CCTGACCAAA ACCGTCAATG
1001  AAAACAAACA AAACGTCGAT GCCAAAGTAA AAGCTGCAGA ATCTGAAATA
1051  GAAAAGTTAA CAACCAAGTT AGCAGACACT GATGCCGCTT TAGCAGATAC
1101  TGATGCCGCT CTGGATGCAA CCACCAACGC CTTGAATAAA TTGGGAGAAA
1151  ATATAACGAC ATTTGCTGAA GAGACTAAGA CAAATATCGT AAAAATTGAT
1201  GAAAAATTAG AAGCCGTGGC TGATACCGTC GACAAGCATG CCGAAGCATT
1251  CAACGATATC GCCGATTCAT TGGATGAAAC CAACACTAAG GCAGACGAAG
1301  CCGTCAAAAC CGCCAATGAA GCCAAACAGA CGGCCGAAGA AACCAAACAA
1351  AACGTCGATG CCAAAGTAAA AGCTGCAGAA ACTGCAGCAG GCAAAGCCGA
1401  AGCTGCCGCT GGCACAGCTA ATACTGCAGC CGACAAGGCC GAAGCTGTCG
1451  CTGCAAAAGT TACCGACATC AAAGCTGATA TCGCTACGAA CAAAGATAAT
1501  ATTGCTAAAA AAGCAAACAG TGCCGACGTG TACACCAGAG AAGAGTCTGA
1551  CAGCAAATTT GTCAGAATTG ATGGTCTGAA CGCTACTACC GAAAAATTGG
1601  ACACACGCTT GGCTTCTGCT GAAAAATCCA TTGCCGATCA CGATACTCGC
1651  CTGAACGGTT TGGATAAAAC AGTGTCAGAC CTGCGCAAAG AAACCCGCCA
1701  AGGCCTTGCA GAACAAGCCG CGCTCTCCGG TCTGTTCCAA CCTTACAACG
1751  TGGGTCTCGA GCACCACCAC CACCACCACT GA

   1  MVAADIGAGL ADALTAPLDH KDKGLQSLTL DQSVRKNEKL KLAAQGAEKT
  51  YGNGDSLNTG KLKNDKVSRF DFIRQIEVDG QLITLESGEF QVYKQSHSAL
 101  TAFQTEQIQD SEHSGKMVAK RQFRIGDIAG EHTSFDKLPE GGRATYRGTA
 151  FGSDDAGGKL TYTIDFAAKQ GNGKIEHLKS PELNVDLAAA DIKPDGKRHA
 201  VISGSVLYNQ AEKGSYSLGI FGGKAQEVAG SAEVKTVNGI RHIGLAAKQL
 251  EGGGGTGSAT NDDDVKKAAT VAIAAAYNNG QEINGFKAGE TIYDIDEDGT
 301  ITKKDATAAD VEADDFKGLG LKKVVTNLTK TVNENKQNVD AKVKAAESEI
 351  EKLTTKLADT DAALADTDAA LDATTNALNK LGENITTFAE ETKTNIVKID
 401  EKLEAVADTV DKHAEAFNDI ADSLDETNTK ADEAVKTANE AKQTAEETKQ
 451  NVDAKVKAAE TAAGKAEAAA GTANTAADKA EAVAAKVTDI KADIATNKDN
 501  IAKKANSADV YTREESDSKF VRIDGLNATT EKLDTRLASA EKSIADHDTR
 551  LNGLDKTVSD LRKETRQGLA EQAALSGLFQ PYNVGLEHHH HHH*
```

ΔG741-983

```
   1  ATGGTCGCCG CCGACATCGG TGCGGGGCTT GCCGATGCAC TAACCGCACC
  51  GCTCGACCAT AAAGACAAAG GTTTGCAGTC TTTGACGCTG GATCAGTCCG
 101  TCAGGAAAAA CGAGAAACTG AAGCTGGCGG CACAAGGTGC GGAAAAAACT
 151  TATGGAAACG GTGACAGCCT CAATACGGGC AAATTGAAGA ACGACAAGGT
 201  CAGCCGTTTC GACTTTATCC GCCAAATCGA AGTGGACGGG CAGCTCATTA
 251  CCTTGGAGAG TGGAGAGTTC CAAGTATACA AACAAAGCCA TTCCGCCTTA
 301  ACCGCCTTTC AGACCGAGCA AATACAAGAT TCGGAGCATT CCGGGAAGAT
 351  GGTTGCGAAA CGCCAGTTCA GAATCGGCGA CATAGCGGGC GAACATACAT
 401  CTTTTGACAA GCTTCCCGAA GGCGGCAGGG CGACATATCG CGGGACGGCG
 451  TTCGGTTCAG ACGATGCCGG CGGAAAACTG ACCTACACCA TAGATTTCGC
 501  CGCCAAGCAG GGAAACGGCA AAATCGAACA TTTGAAATCG CCAGAACTCA
 551  ATGTCGACCT GGCCGCCGCC GATATCAAGC CGGATGGAAA ACGCCATGCC
 601  GTCATCAGCG GTTCCGTCCT TTACAACCAA GCCGAGAAAG GCAGTTACTC
 651  CCTCGGTATC TTTGGCGGAA AAGCCCAGGA AGTTGCCGGC AGCGCGGAAG
 701  TGAAAACCGT AAACGGCATA CGCCATATCG GCCTTGCCGC CAAGCAACTC
 751  GAGGGATCCG GCGGAGGCGG CACTTCTGCG CCCGACTTCA ATGCAGGCGG
 801  TACCGGTATC GGCAGCAACA GCAGAGCAAC AACAGCGAAA TCAGCAGCAG
 851  TATCTTACGC CGGTATCAAG AACGAAATGT GCAAAGACAG AAGCATGCTC
 901  TGTGCCGGTC GGGATGACGT TGCGGTTACA GACAGGGATG CCAAAATCAA
 951  TGCCCCCCCC CCGAATCTGC ATACCGGAGA CTTTCCAAAC CCAAATGACG
1001  CATACAAGAA TTTGATCAAC CTCAAACCTG CAATTGAAGC AGGCTATACA
1051  GGACGCGGGG TAGAGGTAGG TATCGTCGAC ACAGGCGAAT CCGTCGGCAG
1101  CATATCCTTT CCCGAACTGT ATGGCAGAAA AGAACACGGC TATAACGAAA
1151  ATTACAAAAA CTATACGGCG TATATGCGGA AGGAAGCGCC TGAAGACGGA
1201  GGCGGTAAAG ACATTGAAGC TTCTTTCGAC GATGAGGCCG TTATAGAGAC
1251  TGAAGCAAAG CCGACGGATA TCCGCCACGT AAAAGAAATC GGACACATCG
1301  ATTTGGTCTC CCATATTATT GGCGGGCGTT CCGTGGACGG CAGACCTGCA
1351  GGCGGTATTG CGCCCGATGC GACGCTACAC ATAATGAATA CGAATGATGA
1401  AACCAAGAAC GAAATGATGG TTGCAGCCAT CCGCAATGCA TGGGTCAAGC
1451  TGGGCGAACG TGGCGTGCGC ATCGTCAATA ACAGTTTTGG AACAACATCG
1501  AGGGCAGGCA CTGCCGACCT TTTCCAAATA GCCAATTCGG AGGAGCAGTA
```

```
1551  CCGCCAAGCG TTGCTCGACT ATTCCGGCGG TGATAAAACA GACGAGGGTA
1601  TCCGCCTGAT GCAACAGAGC GATTACGGCA ACCTGTCCTA CCACATCCGT
1651  AATAAAAACA TGCTTTTCAT CTTTTCGACA GGCAATGACG CACAAGCTCA
1701  GCCCAACACA TATGCCCTAT TGCCATTTTA TGAAAAAGAC GCTCAAAAAG
1751  GCATTATCAC AGTCGCAGGC GTAGACCGCA GTGGAGAAAA GTTCAAACGG
1801  GAAATGTATG GAGAACCGGG TACAGAACCG CTTGAGTATG GCTCCAACCA
1851  TTGCGGAATT ACTGCCATGT GGTGCCTGTC GGCACCCTAT GAAGCAAGCG
1901  TCCGTTTCAC CCGTACAAAC CCGATTCAAA TTGCCGGAAC ATCCTTTTCC
1951  GCACCCATCG TAACCGGCAC GGCGGCTCTG CTGCTGCAGA AATACCCGTG
2001  GATGAGCAAC GACAACCTGC GTACCACGTT GCTGACGACG GCTCAGGACA
2051  TCGGTGCAGT CGGCGTGGAC AGCAAGTTCG GCTGGGGACT GCTGGATGCG
2101  GGTAAGGCCA TGAACGGACC CGCGTCCTTT CCGTTCGGCG ACTTTACCGC
2151  CGATACGAAA GGTACATCCG ATATTGCCTA CTCCTTCCGT AACGACATTT
2201  CAGGCACGGG CGGCCTGATC AAAAAAGGCG GCAGCCAACT GCAACTGCAC
2251  GGCAACAACA CCTATACGGG CAAAACCATT ATCGAAGGCG GTTCGCTGGT
2301  GTTGTACGGC AACAACAAAT CGGATATGCG CGTCGAAACC AAAGGTGCGC
2351  TGATTTATAA CGGGGCGGCA TCCGGCGGCA GCCTGAACAG CGACGGCATT
2401  GTCTATCTGG CAGATACCGA CCAATCCGGC GCAAACGAAA CCGTACACAT
2451  CAAAGGCAGT CTGCAGCTGG ACGGCAAAGG TACGCTGTAC ACACGTTTGG
2501  GCAAACTGCT GAAAGTGGAC GGTACGGCGA TTATCGGCGG CAAGCTGTAC
2551  ATGTCGGCAC GCGGCAAGGG GGCAGGCTAT CTCAACAGTA CCGGACGACG
2601  TGTTCCCTTC CTGAGTGCCG CCAAAATCGG GCAGGATTAT TCTTTCTTCA
2651  CAAACATCGA AACCGACGGC GGCCTGCTGG CTTCCCTCGA CAGCGTCGAA
2701  AAAACAGCGG GCAGTGAAGG CGACACGCTG TCCTATTATG TCCGTCGCGG
2751  CAATGCGGCA CGGACTGCTT CGGCAGCGGC ACATTCCGCG CCCGCCGGTC
2801  TGAAACACGC CGTAGAACAG GGCGGCAGCA ATCTGGAAAA CCTGATGGTC
2851  GAACTGGATG CCTCCGAATC ATCCGCAACA CCCGAGACGG TTGAAACTGC
2901  GGCAGCCGAC CGCACAGATA TGCCGGGCAT CCGCCCCTAC GGCGCAACTT
2951  TCCGCGCAGC GGCAGCCGTA CAGCATGCGA ATGCCGCCGA CGGTGTACGC
3001  ATCTTCAACA GTCTCGCCGC TACCGTCTAT GCCGACAGTA CCGCCGCCCA
3051  TGCCGATATG CAGGGACGCC GCCTGAAAGC CGTATCGGAC GGGTTGGACC
3101  ACAACGGCAC GGGTCTGCGC GTCATCGCGC AAACCCAACA GGACGGTGGA
3151  ACGTGGGAAC AGGGCGGTGT TGAAGGCAAA ATGCGCGGCA GTACCCAAAC
3201  CGTCGGCATT GCCGCGAAAA CCGGCGAAAA TACGACAGCA GCCGCCACAC
3251  TGGGCATGGG ACGCAGCACA TGGAGCGAAA ACAGTGCAAA TGCAAAAACC
3301  GACAGCATTA GTCTGTTTGC AGGCATACGG CACGATGCGG GCGATATCGG
3351  CTATCTCAAA GGCCTGTTCT CCTACGGACG CTACAAAAAC AGCATCAGCC
3401  GCAGCACCGG TGCGGACGAA CATGCGGAAG GCAGCGTCAA CGGCACGCTG
3451  ATGCAGCTGG GCGCACTGGG CGGTGTCAAC GTTCCGTTTG CCGCAACGGG
3501  AGATTTGACG GTCGAAGGCG GTCTGCGCTA CGACCTGCTC AAACAGGATG
3551  CATTCGCCGA AAAAGGCAGT GCTTTGGGCT GGAGCGGCAA CAGCCTCACT
3601  GAAGGCACGC TGGTCGGACT CGCGGGTCTG AAGCTGTCGC AACCCTTGAG
3651  CGATAAAGCC GTCCTGTTTG CAACGGCGGG CGTGGAACGC GACCTGAACG
3701  GACGCGACTA CACGGTAACG GGCGGCTTTA CCGGCGCGAC TGCAGCAACC
3751  GGCAAGACGG GGGCACGCAA TATGCCGCAC ACCCGTCTGG TTGCCGGCCT
3801  GGGCGCGGAT GTCGAATTCG GCAACGGCTG GAACGGCTTG GCACGTTACA
3851  GCTACGCCGG TTCCAAACAG TACGGCAACC ACAGCGGACG AGTCGGCGTA
3901  GGCTACCGGT TCCTCGAGCA CCACCACCAC CACCACTGA

   1  MVAADIGAGL ADALTAPLDH KDKGLQSLTL DQSVRKNEKL KLAAQGAEKT
  51  YGNGDSLNTG KLKNDKVSRF DFIRQIEVDG QLITLESGEF QVYKQSHSAL
 101  TAFQTEQIQD SEHSGKMVAK RQFRIGDIAG EHTSFDKLPE GGRATYRGTA
 151  FGSDDAGGKL TYTIDFAAKQ GNGKIEHLKS PELNVDLAAA DIKPDGKRHA
 201  VISGSVLYNQ AEKGSYSLGI FGGKAQEVAG SAEVKTVNGI RHIGLAAKQL
 251  EGSGGGGTSA PDFNAGGTGI GSNSRATTAK SAAVSYAGIK NEMCKDRSML
 301  CAGRDDVAVT DRDAKINAPP PNLHTGDFPN PNDAYKNLIN LKPAIEAGYT
 351  GRGVEVGIVD TGESVGSISF PELYGRKEHG YNENYKNYTA YMRKEAPEDG
 401  GGKDIEASFD DEAVIETEAK PTDIRHVKEI GHIDLVSHII GGRSVDGRPA
 451  GGIAPDATLH IMNTNDETKN EMMVAAIRNA WVKLGERGVR IVNNSFGTTS
 501  RAGTADLFQI ANSEEQYRQA LLDYSGGDKT DEGIRLMQQS DYGNLSYHIR
 551  NKNMLFIFST GNDAQAQPNT YALLPFYEKD AQKGIITVAG VDRSGEKFKR
 601  EMYGEPGTEP LEYGSNHCGI TAMWCLSAPY EASVRFTRTN PIQIAGTSFS
 651  APIVTGTAAL LLQKYPWMSN DNLRTTLLTT AQDIGAVGVD SKFGWGLLDA
 701  GKAMNGPASF PFGDFTADTK GTSDIAYSFR NDISGTGGLI KKGGSQLQLH
 751  GNNTYTGKTI IEGGSLVLYG NNKSDMRVET KGALIYNGAA SGGSLNSDGI
 801  VYLADTDQSG ANETVHIKGS LQLDGKGTLY TRLGKLLKVD GTAIIGGKLY
 851  MSARGKGAGY LNSTGRRVPF LSAAKIGQDY SFFTNIETDG GLLASLDSVE
```

```
 901  KTAGSEGDTL SYYVRRGNAA RTASAAAHSA PAGLKHAVEQ GGSNLENLMV
 951  ELDASESSAT PETVETAAAD RTDMPGIRPY GATFRAAAAV QHANAADGVR
1001  IFNSLAATVY ADSTAAHADM QGRRLKAVSD GLDHNGTGLR VIAQTQQDGG
1051  TWEQGGVEGK MRGSTQTVGI AAKTGENTTA AATLGMGRST WSENSANAKT
1101  DSISLFAGIR HDAGDIGYLK GLFSYGRYKN SISRSTGADE HAEGSVNGTL
1151  MQLGALGGVN VPFAATGDLT VEGGLRYDLL KQDAFAEKGS ALGWSGNSLT
1201  EGTLVGLAGL KLSQPLSDKA VLFATAGVER DLNGRDYTVT GGFTGATAAT
1251  GKTGARNMPH TRLVAGLGAD VEFGNGWNGL ARYSYAGSKQ YGNHSGRVGV
1301  GYRFLEHHHH HH*
```

ΔG741–ORF46.1

```
   1  ATGGTCGCCG CCGACATCGG TGCGGGGCTT GCCGATGCAC TAACCGCACC
  51  GCTCGACCAT AAAGACAAAG GTTTGCAGTC TTTGACGCTG GATCAGTCCG
 101  TCAGGAAAAA CGAGAAACTG AAGCTGGCGG CACAAGGTGC GGAAAAAACT
 151  TATGGAAACG GTGACAGCCT CAATACGGGC AAATTGAAGA ACGACAAGGT
 201  CAGCCGTTTC GACTTTATCC GCCAAATCGA AGTGGACGGG CAGCTCATTA
 251  CCTTGGAGAG TGGAGAGTTC CAAGTATACA AACAAAGCCA TTCCGCCTTA
 301  ACCGCCTTTC AGACCGAGCA AATACAAGAT TCGGAGCATT CCGGGAAGAT
 351  GGTTGCGAAA CGCCAGTTCA GAATCGGCGA CATAGCGGGC GAACATACAT
 401  CTTTTGACAA GCTTCCCGAA GGCGGCAGGG CGACATATCG CGGGACGGCG
 451  TTCGGTTCAG ACGATGCCGG CGGAAAACTG ACCTACACCA TAGATTTCGC
 501  CGCCAAGCAG GGAAACGGCA AAATCGAACA TTTGAAATCG CCAGAACTCA
 551  ATGTCGACCT GGCCGCCGCC GATATCAAGC CGGATGGAAA ACGCCATGCC
 601  GTCATCAGCG GTTCCGTCCT TTACAACCAA GCCGAGAAAG GCAGTTACTC
 651  CCTCGGTATC TTTGGCGGAA AAGCCCAGGA AGTTGCCGGC AGCGCGGAAG
 701  TGAAAACCGT AAACGGCATA CGCCATATCG GCCTTGCCGC CAAGCAACTC
 751  GACGGTGGCG GAGGCACTGG ATCCTCAGAT TTGGCAAACG ATTCTTTTAT
 801  CCGGCAGGTT CTCGACCGTC AGCATTTCGA ACCCGACGGG AAATACCACC
 851  TATTCGGCAG CAGGGGGGAA CTTGCCGAGC GCAGCGGCCA TATCGGATTG
 901  GGAAAAATAC AAAGCCATCA GTTGGGCAAC CTGATGATTC AACAGGCGGC
 951  CATTAAAGGA AATATCGGCT ACATTGTCCG CTTTTCCGAT CACGGGCACG
1001  AAGTCCATTC CCCCTTCGAC AACCATGCCT CACATTCCGA TTCTGATGAA
1051  GCCGGTAGTC CCGTTGACGG ATTTAGCCTT TACCGCATCC ATTGGGACGG
1101  ATACGAACAC CATCCCGCCG ACGGCTATGA CGGGCCACAG GGCGGCGGCT
1151  ATCCCGCTCC CAAAGGCGCG AGGGATATAT ACAGCTACGA CATAAAAGGC
1201  GTTGCCCAAA ATATCCGCCT CAACCTGACC GACAACCGCA GCACCGGACA
1251  ACGGCTTGCC GACCGTTTCC ACAATGCCGG TAGTATGCTG ACGCAAGGAG
1301  TAGGCGACGG ATTCAAACGC GCCACCCGAT ACAGCCCCGA GCTGGACAGA
1351  TCGGGCAATG CCGCCGAAGC CTTCAACGGC ACTGCAGATA TCGTTAAAAA
1401  CATCATCGGC GCGGCAGGAG AAATTGTCGG CGCAGGCGAT GCCGTGCAGG
1451  GCATAAGCGA AGGCTCAAAC ATTGCTGTCA TGCACGGCTT GGGTCTGCTT
1501  TCCACCGAAA ACAAGATGGC GCGCATCAAC GATTTGGCAG ATATGGCGCA
1551  ACTCAAAGAC TATGCCGCAG CAGCCATCCG CGATTGGGCA GTCCAAAACC
1601  CCAATGCCGC ACAAGGCATA GAAGCCGTCA GCAATATCTT TATGGCAGCC
1651  ATCCCCATCA AAGGGATTGG AGCTGTTCGG GGAAAATACG GCTTGGGCGG
1701  CATCACGGCA CATCCTATCA AGCGGTCGCA GATGGGCGCG ATCGCATTGC
1751  CGAAAGGGAA ATCCGCCGTC AGCGACAATT TTGCCGATGC GGCATACGCC
1801  AAATACCCGT CCCCTTACCA TTCCCGAAAT ATCCGTTCAA ACTTGGAGCA
1851  GCGTTACGGC AAAGAAAACA TCACCTCCTC AACCGTGCCG CCGTCAAACG
1901  GCAAAAATGT CAAACTGGCA GACCAACGCC ACCCGAAGAC AGGCGTACCG
1951  TTTGACGGTA AAGGGTTTCC GAATTTTGAG AAGCACGTGA AATATGATAC
2001  GCTCGAGCAC CACCACCACC ACCACTGA

   1  MVAADIGAGL ADALTAPLDH KDKGLQSLTL DQSVRKNEKL KLAAQGAEKT
  51  YGNGDSLNTG KLKNDKVSRF DFIRQIEVDG QLITLESGEF QVYKQSHSAL
 101  TAFQTEQIQD SEHSGKMVAK RQFRIGDIAG EHTSFDKLPE GGRATYRGTA
 151  FGSDDAGGKL TYTIDFAAKQ GNGKIEHLKS PELNVDLAAA DIKPDGKRHA
 201  VISGSVLYNQ AEKGSYSLGI FGGKAQEVAG SAEVKTVNGI RHIGLAAKQL
 251  DGGGGTGSSD LANDSFIRQV LDRQHFEPDG KYHLFGSRGE LAERSGHIGL
 301  GKIQSHQLGN LMIQQAAIKG NIGYIVRFSD HGHEVHSPFD NHASHSDSDE
 351  AGSPVDGFSL YRIHWDGYEH HPADGYDGPQ GGGYPAPKGA RDIYSYDIKG
 401  VAQNIRLNLT DNRSTGQRLA DRFHNAGSML TQGVGDGFKR ATRYSPELDR
 451  SGNAAEAFNG TADIVKNIIG AAGEIVGAGD AVQGISEGSN IAVMHGLGLL
 501  STENKMARIN DLADMAQLKD YAAAAIRDWA VQNPNAAQGI EAVSNIFMAA
 551  IPIKGIGAVR GKYGLGGITA HPIKRSQMGA IALPKGKSAV SDNFADAAYA
 601  KYPSPYHSRN IRSNLEQRYG KENITSSTVP PSNGKNVKLA DQRHPKTGVP
 651  FDGKGFPNFE KHVKYDTLEH HHHHH*
```

Example 16 – C-terminal fusions ('hybrids') with 287/ΔG287

According to the invention, hybrids of two proteins A & B may be either NH_2–A–B–COOH or NH_2–B–A–COOH. The effect of this difference was investigated using protein 287 either C-terminal (in '287-His' form) or N-terminal (in ΔG287 form – sequences shown above) to 919, 953 and ORF46.1. A panel of strains was used, including homologous strain 2996. FCA was used as adjuvant:

	287 & 919		287 & 953		287 & ORF46.1	
Strain	*ΔG287-919*	*919-287*	*ΔG287-953*	*953-287*	*ΔG287-46.1*	*46.1-287*
2996	128000	16000	65536	8192	16384	8192
BZ232	256	128	128	<4	<4	<4
1000	2048	<4	<4	<4	<4	<4
MC58	8192	1024	16384	1024	512	128
NGH38	32000	2048	>2048	4096	16384	4096
394/98	4096	32	256	128	128	16
MenA (F6124)	32000	2048	>2048	32	8192	1024
MenC (BZ133)	64000	>8192	>8192	<16	8192	2048

Better bactericidal titres are generally seen with 287 at the N-terminus (in the ΔG form)

When fused to protein 961 [NH_2-ΔG287-961-COOH – sequence shown above], the resulting protein is insoluble and must be denatured and renatured for purification. Following renaturation, around 50% of the protein was found to remain insoluble. The soluble and insoluble proteins were compared, and much better bactericidal titres were obtained with the soluble protein (FCA as adjuvant):

	2996	**BZ232**	**MC58**	**NGH38**	**F6124**	**BZ133**
Soluble	65536	128	4096	>2048	>2048	4096
Insoluble	8192	<4	<4	16	n.d.	n.d.

Titres with the insoluble form were, however, improved by using alum adjuvant instead:

Insoluble	32768	128	4096	>2048	>2048	2048

Example 17 – N-terminal fusions ('hybrids') to 287

Expression of protein 287 as full-length with a C-terminal His-tag, or without its leader peptide but with a C-terminal His-tag, gives fairly low expression levels. Better expression is achieved using a N-terminal GST-fusion.

As an alternative to using GST as an N-terminal fusion partner, 287 was placed at the C-terminus of protein 919 ('919-287'), of protein 953 ('953-287'), and of proteins ORF46.1 ('ORF46.1-287'). In both cases, the leader peptides were deleted, and the hybrids were direct in-frame fusions.

To generate the 953-287 hybrid, the leader peptides of the two proteins were omitted by designing the forward primer downstream from the leader of each sequence; the stop codon sequence was omitted in the 953 reverse primer but included in the 287 reverse primer. For the 953 gene, the 5' and the 3' primers used for amplification included a *Nde*I and a *Bam*HI restriction sites respectively, whereas for the amplification of the 287 gene the 5' and the 3' primers included a *Bam*HI and a *Xho*I restriction sites respectively. In this way a sequential directional cloning of the two genes in pET21b+, using *Nde*I-*Bam*HI (to clone the first gene) and subsequently *Bam*HI-*Xho*I (to clone the second gene) could be achieved.

The 919-287 hybrid was obtained by cloning the sequence coding for the mature portion of 287 into the *Xho*I site at the 3'-end of the 919-His clone in pET21b+. The primers used for amplification of the 287 gene were designed for introducing a *Sal*I restriction site at the 5'- and a *Xho*I site at the 3'- of the PCR fragment. Since the cohesive ends produced by the *Sal*I and *Xho*I restriction enzymes are compatible, the 287 PCR product digested with *Sal*I-*Xho*I could be inserted in the pET21b-919 clone cleaved with *Xho*I.

The ORF46.1-287 hybrid was obtained similarly.

The bactericidal efficacy (homologous strain) of antibodies raised against the hybrid proteins was compared with antibodies raised against simple mixtures of the component antigens:

	Mixture with 287	**Hybrid with 287**
919	32000	16000
953	8192	8192
ORF46.1	128	8192

Data for bactericidal activity against heterologous MenB strains and against serotypes A and C were also obtained for 919-287 and 953-287:

	919		953		ORF46.1	
Strain	*Mixture*	*Hybrid*	*Mixture*	*Hybrid*	*Mixture*	*Hybrid*
MC58	512	1024	512	1024	-	1024
NGH38	1024	2048	2048	4096	-	4096
BZ232	512	128	1024	16	-	-
MenA (F6124)	512	2048	2048	32	-	1024
MenC (C11)	>2048	n.d.	>2048	n.d.	-	n.d.
MenC (BZ133)	>4096	>8192	>4096	<16	-	2048

Hybrids of ORF46.1 and 919 were also constructed. Best results (four-fold higher titre) were achieved with 919 at the N-terminus.

Hybrids 919-519His, ORF97-225His and 225-ORF97His were also tested. These gave moderate ELISA fitres and bactericidal antibody responses.

Example 18 – the leader peptide from ORF4

As shown above, the leader peptide of ORF4 can be fused to the mature sequence of other proteins (*e.g.* proteins 287 and 919). It is able to direct lipidation in *E.coli.*

Example 19 – domains in 564

The protein '564' is very large (2073aa), and it is difficult to clone and express it in complete form. To facilitate expression, the protein has been divided into four domains, as shown in figure 8 (according to the MC58 sequence):

Domain	**A**	**B**	**C**	**D**
Amino Acids	79-360	361-731	732-2044	2045-2073

These domains show the following homologies:

- Domain A shows homology to other bacterial toxins:

```
gb|AAG03431.1|AE004443_9probable hemagglutinin [Pseudomonas aeruginosa] (38%)
gb|AAC31981.1|(139897)  HecA [Pectobacterium chrysanthemi] (45%)
emb|CAA36409.1|(X52156) filamentous hemagglutinin [Bordetella pertussis] (31%)
gb|AAC79757.1|(AF057695)large supernatant protein1 [Haemophilus ducreyi] (26%)
gb|AAA25657.1|(M30186)  HpmA precursor [Proteus mirabilis] (29%)
```

- Domain B shows no homology, and is specific to 564.

- Domain C shows homology to:

```
gb|AAF84995.1|AE004032  HA-like secreted protein [Xylella fastidiosa] (33%)
gb|AAG05850.1|AE004673  hypothetical protein [Pseudomonas aeruginosa] (27%)
gb|AAF68414.1AF237928   putative FHA [Pasteurella multocisida] (23%)
gb|AAC79757.1|(AF057695)large supernatant protein1 [Haemophilus ducreyi] (23%)
pir||S21010             FHA B precursor [Bordetella pertussis] (20%)
```

• Domain D shows homology to other bacterial toxins:

gb|AAF84995.1|AE004032_14 HA-like secreted protein [Xylella fastidiosa] (29%)

Using the MC58 strain sequence, good intracellular expression of 564ab was obtained in the form of GST-fusions (no purification) and his-tagged protein; this domain-pair was also expressed as a lipoprotein, which showed moderate expression in the outer membrane/ supernatant fraction.

The b domain showed moderate intracellular expression when expressed as a his-tagged product (no purification), and good expression as a GST-fusion.

The c domain showed good intracellular expression as a GST-fusion, but was insoluble. The d domain showed moderate intracellular expression as a his-tagged product (no purification). The cd protein domain-pair showed moderate intracellular expression (no purification) as a GST-fusion.

Good bactericidal assay titres were observed using the c domain and the bc pair.

Example 20 – the 919 leader peptide

The 20mer leader peptide from 919 is discussed in example 1 above:

MKKYLFRAAL YGIAAAILAA

As shown in example 1, deletion of this leader improves heterologous expression, as does substitution with the ORF4 leader peptide. The influence of the 919 leader on expression was investigated by fusing the coding sequence to the *PhoC* reporter gene from *Morganella morganii* [Thaller *et al.* (1994) *Microbiology* 140:1341-1350]. The construct was cloned in the pET21-b plasmid between the *Nde*I and *Xho*I sites (Figure 9):

```
  1  MKKYLFRAAL YGIAAAILAA AIPAGNDATT KPDLYYLKNE QAIDSLKLLP
 51  PPPEVGSIQF LNDQAMYEKG RMLRNTERGK QAQADADLAA GGVATAFSGA
101  FGYPITEKDS PELYKLLTNM IEDAGDLATR SAKEHYMRIR PFAFYGTETC
151  NTKDQKKLST NGSYPSGHTS IGWATALVLA EVNPANQDAI LERGYQLGQS
201  RVICGYHWQS DVDAARIVGS AAVATLHSDP AFQAQLAKAK QEFAQKSQK*
```

The level of expression of PhoC from this plasmid is >200-fold lower than that found for the same construct but containing the native PhoC signal peptide. The same result was obtained even after substitution of the T7 promoter with the *E.coli* Plac promoter. This means that the influence of the 919 leader sequence on expression does not depend on the promoter used.

In order to investigate if the results observed were due to some peculiarity of the 919 signal peptide nucleotide sequence (secondary structure formation, sensitivity to RNAases, *etc.*) or

to protein instability induced by the presence of this signal peptide, a number of mutants were generated. The approach used was a substitution of nucleotides of the 919 signal peptide sequence by cloning synthetic linkers containing degenerate codons. In this way, mutants were obtained with nucleotide and/or amino acid substitutions.

Two different linkers were used, designed to produce mutations in two different regions of the 919 signal peptide sequence, in the first 19 base pairs (L1) and between bases 20-36 (S1).

```
L1: 5' T ATG AAa/g TAc/t c/tTN TTt/c a/cGC GCC GCC CTG TAC GGC ATC GCC GCC
GCC ATC CTC GCC GCC GCG ATC CC 3'
S1: 5' T ATG AAA AAA TAC CTA TTC CGa/g GCN GCN c/tTa/g TAc/t GGc/g ATC GCC
GCC GCC ATC CTC GCC GCC GCG ATC CC 3'
```

The alignment of some of the mutants obtained is given below.

```
L1 mutants:
9L1-a    ATGAAGAAGTACCTTTTCAGCGCCGCC~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
9L1-e    ATGAAAAAATACTTTTTCCGCGCCGCC~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
9L1-d    ATGAAAAAATACTTTTTCCGCGCCGCC~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
9L1-f    ATGAAAAAATATCTCTTTAGCGCCGCCCTGTACGGCATCGCCGCCGCCATCCTCGCCGCC
919sp    ATGAAAAAATACCTATTCCGCGCCGCCCTGTACGGCATCGCCGCCGCCATCCTCGCCGCC

9L1a     MKKYLFSAA~~~~~~~~~~~
9L1e     MKKYFFRAA~~~~~~~~~~~
9L1d     MKKYFFRAA~~~~~~~~~~~
9L1f     MKKYLFSAALYGIAAAILAA
919sp    MKKYLFRAALYGIAAAILAA (i.e. native signal peptide)

S1 mutants:
9S1-e   ATGAAAAAATACCTATTC.................ATCGCCGCCGCCATCCTCGCCGCC
9S1-c   ATGAAAAAATACCTATTCCGAGCTGCCCAATACGGCATCGCCGCCGCCATCCTCGCCGCC
9S1-b   ATGAAAAAATACCTATTCCGGGCCGCCCAATACGGCATCGCCGCCGCCATCCTCGCCGCC
9S1-i   ATGAAAAAATACCTATTCCGGGCGGCTTTGTACGGGATCGCCGCCGCCATCCTCGCCGCC
919sp   ATGAAAAAATACCTATTCCGCGCCGCCCTGTACGGCATCGCCGCCGCCATCCTCGCCGCC

9S1e   MKKYLF......IAAAILAA
9S1c   MKKYLFRAAQYGIAAAILAA
9S1b   MKKYLFRAAQYGIAAAILAA
9S1i   MKKYLFRAALYGIAAAILAA
919sp  MKKYLFRAALYGIAAAILAA
```

As shown in the sequences alignments, most of the mutants analysed contain in-frame deletions which were unexpectedly produced by the host cells.

Selection of the mutants was performed by transforming *E. coli* BL21(DE3) cells with DNA prepared from a mixture of L1 and S1 mutated clones. Single transformants were screened for high PhoC activity by streaking them onto LB plates containing 100 µg/ml ampicillin, 50µg/ml methyl green, 1 mg/ml PDP (phenolphthaleindiphosphate). On this medium PhoC-producing cells become green (Figure 10).

A quantitative analysis of PhoC produced by these mutants was carried out in liquid medium using pNPP as a substrate for PhoC activity. The specific activities measured in cell extracts and supernatants of mutants grown in liquid medium for 0, 30, 90, 180 min. were:

CELL EXTRACTS

	0	30	90	180
control	0,00	0,00	0,00	0,00
9phoC	1,11	1,11	3,33	4,44
9S1e	102,12	111,00	149,85	172,05
9L1a	206,46	111,00	94,35	83,25
9L1d	5,11	4,77	4,00	3,11
9L1f	27,75	94,35	82,14	36,63
9S1b	156,51	111,00	72,15	28,86
9S1c	72,15	33,30	21,09	14,43
9S1i	156,51	83,25	55,50	26,64
phoCwt	194,25	180,93	149,85	142,08

SUPERNATANTS

	0	30	90	180
control	0,00	0,00	0,00	0,00
9phoC	0,33	0,00	0,00	0,00
9S1e	0,11	0,22	0,44	0,89
9L1a	4,88	5,99	5,99	7,22
9L1d	0,11	0,11	0,11	0,11
9L1f	0,11	0,22	0,11	0,11
9S1b	1,44	1,44	1,44	1,67
9S1c	0,44	0,78	0,56	0,67
9S1i	0,22	0,44	0,22	0,78
phoCwt	34,41	43,29	87,69	177,60

Some of the mutants produce high amounts of PhoC and in particular, mutant 9L1a can secrete PhoC in the culture medium. This is noteworthy since the signal peptide sequence of this mutant is only 9 amino acids long. This is the shortest signal peptide described to date.

Example 21 – C-terminal deletions of Maf-related proteins

MafB-related proteins include 730, ORF46 and ORF29.

The 730 protein from MC58 has the following sequence:

```
  1  VKPLRRLTNL LAACAVAAAA LIQPALAADL AQDPFITDNA QRQHYEPGGK
 51  YHLFGDPRGS VSDRTGKINV IQDYTHQMGN LLIQQANING TIGYHTRFSG
101  HGHEEHAPFD NHAADSASEE KGNVDEGFTV YRLNWEGHEH HPADAYDGPK
151  GGNYPKPTGA RDEYTYHVNG TARSIKLNPT DTRSIRQRIS DNYSNLGSNF
201  SDRADEANRK MFEHNAKLDR WGNSMEFING VAAGALNPFI SAGEALGIGD
251  ILYGTRYAID KAAMRNIAPL PAEGKFAVIG GLGSVAGFEK NTREAVDRWI
301  QENPNAAETV EAVFNVAAAA KVAKLAKAAK PGKAAVSGDF ADSYKKKLAL
```

```
351  SDSARQLYQN AKYREALDIH YEDLIRRKTD GSSKFINGRE IDAVTNDALI
401  QAKRTISAID KPKNFLNQKN RKQIKATIEA ANQQGKRAEF WFKYGVHSQV
451  KSYIESKGGI VKTGLGD*
```

The leader peptide is underlined.

730 shows similar features to ORF46 (see example 8 above):

- as for Orf46, the conservation of the 730 sequence among MenB, MenA and gonococcus is high (>80%) only for the N-terminal portion. The C-terminus, from ~340, is highly divergent.
- its predicted secondary structure contains a hydrophobic segment spanning the central region of the molecule (aa. 227-247).
- expression of the full-length gene in *E. coli* gives very low yields of protein. Expression from tagged or untagged constructs where the signal peptide sequence has been omitted has a toxic effect on the host cells. In other words, the presence of the full-length mature protein in the cytoplasm is highly toxic for the host cell while its translocation to the periplasm (mediated by the signal peptide) has no detectable effect on cell viability. This "intracellular toxicity" of 730 is particularly high since clones for expression of the leaderless 730 can only be obtained at very low frequency using a *recA* genetic background (*E. coli* strains: HB101 for cloning; HMS174(DE3) for expression).

To overcome this toxicity, a similar approach was used for 730 as described in example 8 for ORF46. Four C-terminal truncated forms were obtained, each of which is well expressed. All were obtained from intracellular expression of His-tagged leaderless 730.

Form A consists of the N-terminal hydrophilic region of the mature protein (aa. 28-226). This was purified as a soluble His-tagged product, having a higher-than-expected MW.

Form B extends to the end of the region conserved between serogroups (aa. 28-340). This was purified as an insoluble His-tagged product.

The C-terminal truncated forms named C1 and C2 were obtained after screening for clones expressing high levels of 730-His clones in strain HMS174(DE3). Briefly, the pET21b plasmid containing the His-tagged sequence coding for the full-length mature 730 protein was used to transform the *recA* strain HMS174(DE3). Transformants were obtained at low frequency which showed two phenotypes: large colonies and very small colonies. Several large and small colonies were analysed for expression of the 730-His clone. Only cells from large colonies over-expressed a protein recognised by anti-730A antibodies. However the

protein over-expressed in different clones showed differences in molecular mass. Sequencing of two of the clones revealed that in both cases integration of an *E. coli* IS sequence had occurred within the sequence coding for the C terminal region of 730. The two integration events have produced in-frame fusion with 1 additional codon in the case of C1, and 12 additional codons in the case of C2 (Figure 11). The resulting "mutant" forms of 730 have the following sequences:

730-C1 (due to an IS1 insertion - figure 11A)

```
   1  MADLAQDPFI TDNAQRQHYE PGGKYHLFGD PRGSVSDRTG KINVIQDYTH
  51  QMGNLLIQQA NINGTIGYHT RFSGHGHEEH APFDNHAADS ASEEKGNVDE
 101  GFTVYRLNWE GHEHHPADAY DGPKGGNYPK PTGARDEYTY HVNGTARSIK
 151  LNPTDTRSIR QRISDNYSNL GSNFSDRADE ANRKMFEHNA KLDRWGNSME
 201  FINGVAAGAL NPFISAGEAL GIGDILYGTR YAIDKAAMRN IAPLPAEGKF
 251  AVIGGLGSVA GFEKNTREAV DRWIQENPNA AETVEAVFNV AAAAKVAKLA
 301  KAAKPGKAAV SGDFADSYKK KLALSDSARQ LYQNAKYREA LDIHYEDLIR
 351  RKTDGSSKFI NGREIDAVTN DALIQAR*
```

The additional amino acid produced by the insertion is underlined.

730-C2 (due to an IS5 insertion - Figure 11B)

```
   1  MADLAQDPFI TDNAQRQHYE PGGKYHLFGD PRGSVSDRTG KINVIQDYTH
  51  QMGNLLIQQA NINGTIGYHT RFSGHGHEEH APFDNHAADS ASEEKGNVDE
 101  GFTVYRLNWE GHEHHPADAY DGPKGGNYPK PTGARDEYTY HVNGTARSIK
 151  LNPTDTRSIR QRISDNYSNL GSNFSDRADE ANRKMFEHNA KLDRWGNSME
 201  FINGVAAGAL NPFISAGEAL GIGDILYGTR YAIDKAAMRN IAPLPAEGKF
 251  AVIGGLGSVA GFEKNTREAV DRWIQENPNA AETVEAVFNV AAAAKVAKLA
 301  KAAKPGKAAV SGDFADSYKK KLALSDSARQ LYQNAKYREA LGKVRISGEI
 351  LLG*
```

The additional amino acids produced by the insertion are underlined.

In conclusion, intracellular expression of the 730-C1 form gives very high level of protein and has no toxic effect on the host cells, whereas the presence of the native C-terminus is toxic. These data suggest that the "intracellular toxicity" of 730 is associated with the C-terminal 65 amino acids of the protein.

Equivalent truncation of ORF29 to the first 231 or 368 amino acids has been performed, using expression with or without the leader peptide (amino acids 1-26; deletion gives cytoplasmic expression) and with or without a His-tag.

Example 22 – domains in 961

As described in example 9 above, the GST-fusion of 961 was the best-expressed in *E.coli.* To improve expression, the protein was divided into domains (figure 12).

The domains of 961 were designed on the basis of YadA (an adhesin produced by *Yersinia* which has been demonstrated to be an adhesin localized on the bacterial surface that forms

oligomers that generate surface projection [Hoiczyk *et al.* (2000) *EMBO J* 19:5989-99]) and are: leader peptide, head domain, coiled-coil region (stalk), and membrane anchor domain.

These domains were expressed with or without the leader peptide, and optionally fused either to C-terminal His-tag or to N-terminal GST. *E.coli* clones expressing different domains of 961 were analyzed by SDS-PAGE and western blot for the production and localization of the expressed protein, from over-night (o/n) culture or after 3 hours induction with IPTG. The results were:

	Total lysate (Western Blot)	**Periplasm (Western Blot)**	**Supernatant (Western Blot)**	**OMV SDS-PAGE**
961 (o/n)	-	-	-	
961 (IPTG)	+/-	-	-	
961-L (o/n)	+	-	-	+
961-L (IPTG)	+	-	-	+
961c-L (o/n)	-	-	-	
961c-L (IPTG)	+	+	+	
961Δ_1-L (o/n)	-	-	-	
961Δ_1-L (IPTG)	+	-	-	+

The results show that in *E.coli:*

- 961-L is highly expressed and localized on the outer membrane. By western blot analysis two specific bands have been detected: one at ~45kDa (the predicted molecular weight) and one at ~180kDa, indicating that 961-L can form oligomers. Additionally, these aggregates are more expressed in the over-night culture (without IPTG induction). OMV preparations of this clone were used to immunize mice and serum was obtained. Using overnight culture (predominantly by oligomeric form) the serum was bactericidal; the IPTG-induced culture (predominantly monomeric) was not bactericidal.

- 961Δ_1-L (with a partial deletion in the anchor region) is highly expressed and localized on the outer membrane, but does not form oligomers;

- the 961c-L (without the anchor region) is produced in soluble form and exported in the supernatant.

Titres in ELISA and in the serum bactericidal assay using His-fusions were as follows:

	ELISA	**Bactericidal**
961a (aa 24-268)	24397	4096

961b (aa 269-405)	7763	64
961c-L	29770	8192
961c (2996)	30774	>65536
961c (MC58)	33437	16384
961d	26069	>65536

E.coli clones expressing different forms of 961 (961, 961-L, 961Δ_1-L and 961c-L) were used to investigate if the 961 is an adhesin (*c.f.* YadA). An adhesion assay was performed using (a) the human epithelial cells and (b) *E.coli* clones after either over-night culture or three hours IPTG induction. 961-L grown over-night (961Δ_1-L) and IPTG-induced 961c-L (the clones expressing protein on surface) adhere to human epithelial cells.

961c was also used in hybrid proteins (see above). As 961 and its domain variants direct efficient expression, they are ideally suited as the N-terminal portion of a hybrid protein.

Example 23 – further hybrids

Further hybrid proteins of the invention are shown below (see also Figure 14). These are advantageous when compared to the individual proteins:

ORF46.1-741

```
   1  ATGTCAGATT TGGCAAACGA TTCTTTTATC CGGCAGGTTC TCGACCGTCA
  51  GCATTTCGAA CCCGACGGGA AATACCACCT ATTCGGCAGC AGGGGGGAAC
 101  TTGCCGAGCG CAGCGGCCAT ATCGGATTGG GAAAAATACA AAGCCATCAG
 151  TTGGGCAACC TGATGATTCA ACAGGCGGCC ATTAAAGGAA ATATCGGCTA
 201  CATTGTCCGC TTTTCCGATC ACGGGCACGA AGTCCATTCC CCCTTCGACA
 251  ACCATGCCTC ACATTCCGAT TCTGATGAAG CCGGTAGTCC CGTTGACGGA
 301  TTTAGCCTTT ACCGCATCCA TTGGGACGGA TACGAACACC ATCCCGCCGA
 351  CGGCTATGAC GGGCCACAGG GCGGCGGCTA TCCCGCTCCC AAAGGCGCGA
 401  GGGATATATA CAGCTACGAC ATAAAAGGCG TTGCCCAAAA TATCCGCCTC
 451  AACCTGACCG ACAACCGCAG CACCGGACAA CGGCTTGCCG ACCGTTTCCA
 501  CAATGCCGGT AGTATGCTGA CGCAAGGAGT AGGCGACGGA TTCAAACGCG
 551  CCACCCGATA CAGCCCCGAG CTGGACAGAT CGGGCAATGC CGCCGAAGCC
 601  TTCAACGGCA CTGCAGATAT CGTTAAAAAC ATCATCGGCG CGGCAGGAGA
 651  AATTGTCGGC GCAGGCGATG CCGTGCAGGG CATAAGCGAA GGCTCAAACA
 701  TTGCTGTCAT GCACGGCTTG GGTCTGCTTT CCACCGAAAA CAAGATGGCG
 751  CGCATCAACG ATTTGGCAGA TATGGCGCAA CTCAAAGACT ATGCCGCAGC
 801  AGCCATCCGC GATTGGGCAG TCCAAAACCC CAATGCCGCA CAAGGCATAG
 851  AAGCCGTCAG CAATATCTTT ATGGCAGCCA TCCCCATCAA AGGGATTGGA
 901  GCTGTTCGGG GAAAATACGG CTTGGGCGGC ATCACGGCAC ATCCTATCAA
 951  GCGGTCGCAG ATGGGCGCGA TCGCATTGCC GAAAGGGAAA TCCGCCGTCA
1001  GCGACAATTT TGCCGATGCG GCATACGCCA AATACCCGTC CCCTTACCAT
1051  TCCCGAAATA TCCGTTCAAA CTTGGAGCAG CGTTACGGCA AAGAAAACAT
1101  CACCTCCTCA ACCGTGCCGC CGTCAAACGG CAAAAATGTC AAACTGGCAG
1151  ACCAACGCCA CCCGAAGACA GGCGTACCGT TTGACGGTAA AGGGTTTCCG
1201  AATTTTGAGA AGCACGTGAA ATATGATACG GGATCCGGAG GGGGTGGTGT
1251  CGCCGCCGAC ATCGGTGCGG GGCTTGCCGA TGCACTAACC GCACCGCTCG
1301  ACCATAAAGA CAAAGGTTTG CAGTCTTTGA CGCTGGATCA GTCCGTCAGG
1351  AAAAACGAGA AACTGAAGCT GGCGGCACAA GGTGCGGAAA AAACTTATGG
1401  AAACGGTGAC AGCCTCAATA CGGGCAAATT GAAGAACGAC AAGGTCAGCC
1451  GTTTCGACTT TATCCGCCAA ATCGAAGTGG ACGGGCAGCT CATTACCTTG
1501  GAGAGTGGAG AGTTCCAAGT ATACAAACAA AGCCATTCCG CCTTAACCGC
1551  CTTTCAGACC GAGCAAATAC AAGATTCGGA GCATTCCGGG AAGATGGTTG
1601  CGAAACGCCA GTTCAGAATC GGCGACATAG CGGGCGAACA TACATCTTTT
```

```
1651   GACAAGCTTC CCGAAGGCGG CAGGGCGACA TATCGCGGGA CGGCGTTCGG
1701   TTCAGACGAT GCCGGCGGAA AACTGACCTA CACCATAGAT TTCGCCGCCA
1751   AGCAGGGAAA CGGCAAAATC GAACATTTGA AATCGCCAGA ACTCAATGTC
1801   GACCTGGCCG CCGCCGATAT CAAGCCGGAT GGAAAACGCC ATGCCGTCAT
1851   CAGCGGTTCC GTCCTTTACA ACCAAGCCGA GAAAGGCAGT TACTCCCTCG
1901   GTATCTTTGG CGGAAAAGCC CAGGAAGTTG CCGGCAGCGC GGAAGTGAAA
1951   ACCGTAAACG GCATACGCCA TATCGGCCTT GCCGCCAAGC AACTCGAGCA
2001   CCACCACCAC CACCACTGA

   1   MSDLANDSFI RQVLDRQHFE PDGKYHLFGS RGELAERSGH IGLGKIQSHQ
  51   LGNLMIQQAA IKGNIGYIVR FSDHGHEVHS PFDNHASHSD SDEAGSPVDG
 101   FSLYRIHWDG YEHHPADGYD GPQGGGYPAP KGARDIYSYD IKGVAQNIRL
 151   NLTDNRSTGQ RLADRFHNAG SMLTQGVGDG FKRATRYSPE LDRSGNAAEA
 201   FNGTADIVKN IIGAAGEIVG AGDAVQGISE GSNIAVMHGL GLLSTENKMA
 251   RINDLADMAQ LKDYAAAAIR DWAVQNPNAA QGIEAVSNIF MAAIPIKGIG
 301   AVRGKYGLGG ITAHPIKRSQ MGAIALPKGK SAVSDNFADA AYAKYPSPYH
 351   SRNIRSNLEQ RYGKENITSS TVPPSNGKNV KLADQRHPKT GVPFDGKGFP
 401   NFEKHVKYDT GSGGGGVAAD IGAGLADALT APLDHKDKGL QSLTLDQSVR
 451   KNEKLKLAAQ GAEKTYGNGD SLNTGKLKND KVSRFDFIRQ IEVDGQLITL
 501   ESGEFQVYKQ SHSALTAFQT EQIQDSEHSG KMVAKRQFRI GDIAGEHTSF
 551   DKLPEGGRAT YRGTAFGSDD AGGKLTYTID FAAKQGNGKI EHLKSPELNV
 601   DLAAADIKPD GKRHAVISGS VLYNQAEKGS YSLGIFGGKA QEVAGSAEVK
 651   TVNGIRHIGL AAKQLEHHHH HH*
```

ORF46.1-961

```
   1   ATGTCAGATT TGGCAAACGA TTCTTTTATC CGGCAGGTTC TCGACCGTCA
  51   GCATTTCGAA CCCGACGGGA AATACCACCT ATTCGGCAGC AGGGGGGAAC
 101   TTGCCGAGCG CAGCGGCCAT ATCGGATTGG GAAAAATACA AAGCCATCAG
 151   TTGGGCAACC TGATGATTCA ACAGGCGGCC ATTAAAGGAA ATATCGGCTA
 201   CATTGTCCGC TTTTCCGATC ACGGGCACGA AGTCCATTCC CCCTTCGACA
 251   ACCATGCCTC ACATTCCGAT TCTGATGAAG CCGGTAGTCC CGTTGACGGA
 301   TTTAGCCTTT ACCGCATCCA TTGGGACGGA TACGAACACC ATCCCGCCGA
 351   CGGCTATGAC GGGCCACAGG GCGGCGGCTA TCCCGCTCCC AAAGGCGCGA
 401   GGGATATATA CAGCTACGAC ATAAAAGGCG TTGCCCAAAA TATCCGCCTC
 451   AACCTGACCG ACAACCGCAG CACCGGACAA CGGCTTGCCG ACCGTTTCCA
 501   CAATGCCGGT AGTATGCTGA CGCAAGGAGT AGGCGACGGA TTCAAACGCG
 551   CCACCCGATA CAGCCCCGAG CTGGACAGAT CGGGCAATGC CGCCGAAGCC
 601   TTCAACGGCA CTGCAGATAT CGTTAAAAAC ATCATCGGCG CGGCAGGAGA
 651   AATTGTCGGC GCAGGCGATG CCGTGCAGGG CATAAGCGAA GGCTCAAACA
 701   TTGCTGTCAT GCACGGCTTG GGTCTGCTTT CCACCGAAAA CAAGATGGCG
 751   CGCATCAACG ATTTGGCAGA TATGGCGCAA CTCAAAGACT ATGCCGCAGC
 801   AGCCATCCGC GATTGGGCAG TCCAAAACCC CAATGCCGCA CAAGGCATAG
 851   AAGCCGTCAG CAATATCTTT ATGGCAGCCA TCCCCATCAA AGGGATTGGA
 901   GCTGTTCGGG GAAAATACGG CTTGGGCGGC ATCACGGCAC ATCCTATCAA
 951   GCGGTCGCAG ATGGGCGCGA TCGCATTGCC GAAAGGGAAA TCCGCCGTCA
1001   GCGACAATTT TGCCGATGCG GCATACGCCA AATACCCGTC CCCTTACCAT
1051   TCCCGAAATA TCCGTTCAAA CTTGGAGCAG CGTTACGGCA AAGAAAACAT
1101   CACCTCCTCA ACCGTGCCGC CGTCAAACGG CAAAAATGTC AAACTGGCAG
1151   ACCAACGCCA CCCGAAGACA GGCGTACCGT TTGACGGTAA AGGGTTTCCG
1201   AATTTTGAGA AGCACGTGAA ATATGATACG GGATCCGGAG GAGGAGGAGC
1251   CACAAACGAC GACGATGTTA AAAAAGCTGC CACTGTGGCC ATTGCTGCTG
1301   CCTACAACAA TGGCCAAGAA ATCAACGGTT TCAAAGCTGG AGAGACCATC
1351   TACGACATTG ATGAAGACGG CACAATTACC AAAAAAGACG CAACTGCAGC
1401   CGATGTTGAA GCCGACGACT TTAAAGGTCT GGGTCTGAAA AAAGTCGTGA
1451   CTAACCTGAC CAAAACCGTC AATGAAAACA AACAAAACGT CGATGCCAAA
1501   GTAAAAGCTG CAGAATCTGA AATAGAAAAG TTAACAACCA AGTTAGCAGA
1551   CACTGATGCC GCTTTAGCAG ATACTGATGC CGCTCTGGAT GCAACCACCA
1601   ACGCCTTGAA TAAATTGGGA GAAAATATAA CGACATTTGC TGAAGAGACT
1651   AAGACAAATA TCGTAAAAAT TGATGAAAAA TTAGAAGCCG TGGCTGATAC
1701   CGTCGACAAG CATGCCGAAG CATTCAACGA TATCGCCGAT TCATTGGATG
1751   AAACCAACAC TAAGGCAGAC GAAGCCGTCA AAACCGCCAA TGAAGCCAAA
1801   CAGACGGCCG AAGAAACCAA ACAAAACGTC GATGCCAAAG TAAAAGCTGC
1851   AGAAACTGCA GCAGGCAAAG CCGAAGCTGC CGCTGGCACA GCTAATACTG
1901   CAGCCGACAA GGCCGAAGCT GTCGCTGCAA AAGTTACCGA CATCAAAGCT
1951   GATATCGCTA CGAACAAAGA TAATATTGCT AAAAAAGCAA ACAGTGCCGA
2001   CGTGTACACC AGAGAAGAGT CTGACAGCAA ATTTGTCAGA ATTGATGGTC
```

```
2051   TGAACGCTAC TACCGAAAAA TTGGACACAC GCTTGGCTTC TGCTGAAAAA
2101   TCCATTGCCG ATCACGATAC TCGCCTGAAC GGTTTGGATA AAACAGTGTC
2151   AGACCTGCGC AAAGAAACCC GCCAAGGCCT TGCAGAACAA GCCGCGCTCT
2201   CCGGTCTGTT CCAACCTTAC AACGTGGGTC GGTTCAATGT AACGGCTGCA
2251   GTCGGCGGCT ACAAATCCGA ATCGGCAGTC GCCATCGGTA CCGGCTTCCG
2301   CTTTACCGAA AACTTTGCCG CCAAAGCAGG CGTGGCAGTC GGCACTTCGT
2351   CCGGTTCTTC CGCAGCCTAC CATGTCGGCG TCAATTACGA GTGGCTCGAG
2401   CACCACCACC ACCACCACTG A

   1   MSDLANDSFI RQVLDRQHFE PDGKYHLFGS RGELAERSGH IGLGKIQSHQ
  51   LGNLMIQQAA IKGNIGYIVR FSDHGHEVHS PFDNHASHSD SDEAGSPVDG
 101   FSLYRIHWDG YEHHPADGYD GPQGGGYPAP KGARDIYSYD IKGVAQNIRL
 151   NLTDNRSTGQ RLADRFHNAG SMLTQGVGDG FKRATRYSPE LDRSGNAAEA
 201   FNGTADIVKN IIGAAGEIVG AGDAVQGISE GSNIAVMHGL GLLSTENKMA
 251   RINDLADMAQ LKDYAAAAIR DWAVQNPNAA QGIEAVSNIF MAAIPIKGIG
 301   AVRGKYGLGG ITAHPIKRSQ MGAIALPKGK SAVSDNFADA AYAKYPSPYH
 351   SRNIRSNLEQ RYGKENITSS TVPPSNGKNV KLADQRHPKT GVPFDGKGFP
 401   NFEKHVKYDT GSGGGGATND DDVKKAATVA IAAAYNNGQE INGFKAGETI
 451   YDIDEDGTIT KKDATAADVE ADDFKGLGLK KVVTNLTKTV NENKQNVDAK
 501   VKAAESEIEK LTTKLADTDA ALADTDAALD ATTNALNKLG ENITTFAEET
 551   KTNIVKIDEK LEAVADTVDK HAEAFNDIAD SLDETNTKAD EAVKTANEAK
 601   QTAEETKQNV DAKVKAAETA AGKAEAAAGT ANTAADKAEA VAAKVTDIKA
 651   DIATNKDNIA KKANSADVYT REESDSKFVR IDGLNATTEK LDTRLASAEK
 701   SIADHDTRLN GLDKTVSDLR KETRQGLAEQ AALSGLFQPY NVGRFNVTAA
 751   VGGYKSESAV AIGTGFRFTE NFAAKAGVAV GTSSGSSAAY HVGVNYEWLE
 801   HHHHHH*
```

ORF46.1-961c

```
   1   ATGTCAGATT TGGCAAACGA TTCTTTTATC CGGCAGGTTC TCGACCGTCA
  51   GCATTTCGAA CCCGACGGGA AATACCACCT ATTCGGCAGC AGGGGGGAAC
 101   TTGCCGAGCG CAGCGGCCAT ATCGGATTGG GAAAAATACA AAGCCATCAG
 151   TTGGGCAACC TGATGATTCA ACAGGCGGCC ATTAAAGGAA ATATCGGCTA
 201   CATTGTCCGC TTTTCCGATC ACGGGCACGA AGTCCATTCC CCCTTCGACA
 251   ACCATGCCTC ACATTCCGAT TCTGATGAAG CCGGTAGTCC CGTTGACGGA
 301   TTTAGCCTTT ACCGCATCCA TTGGGACGGA TACGAACACC ATCCCGCCGA
 351   CGGCTATGAC GGGCCACAGG GCGGCGGCTA TCCCGCTCCC AAAGGCGCGA
 401   GGGATATATA CAGCTACGAC ATAAAAGGCG TTGCCCAAAA TATCCGCCTC
 451   AACCTGACCG ACAACCGCAG CACCGGACAA CGGCTTGCCG ACCGTTTCCA
 501   CAATGCCGGT AGTATGCTGA CGCAAGGAGT AGGCGACGGA TTCAAACGCG
 551   CCACCCGATA CAGCCCCGAG CTGGACAGAT CGGGCAATGC CGCCGAAGCC
 601   TTCAACGGCA CTGCAGATAT CGTTAAAAAC ATCATCGGCG CGGCAGGAGA
 651   AATTGTCGGC GCAGGCGATG CCGTGCAGGG CATAAGCGAA GGCTCAAACA
 701   TTGCTGTCAT GCACGGCTTG GGTCTGCTTT CCACCGAAAA CAAGATGGCG
 751   CGCATCAACG ATTTGGCAGA TATGGCGCAA CTCAAAGACT ATGCCGCAGC
 801   AGCCATCCGC GATTGGGCAG TCCAAAACCC CAATGCCGCA CAAGGCATAG
 851   AAGCCGTCAG CAATATCTTT ATGGCAGCCA TCCCCATCAA AGGGATTGGA
 901   GCTGTTCGGG GAAAATACGG CTTGGGCGGC ATCACGGCAC ATCCTATCAA
 951   GCGGTCGCAG ATGGGCGCGA TCGCATTGCC GAAAGGGAAA TCCGCCGTCA
1001   GCGACAATTT TGCCGATGCG GCATACGCCA AATACCCGTC CCCTTACCAT
1051   TCCCGAAATA TCCGTTCAAA CTTGGAGCAG CGTTACGGCA AAGAAAACAT
1101   CACCTCCTCA ACCGTGCCGC CGTCAAACGG CAAAAATGTC AAACTGGCAG
1151   ACCAACGCCA CCCGAAGACA GGCGTACCGT TTGACGGTAA AGGGTTTCCG
1201   AATTTTGAGA AGCACGTGAA ATATGATACG GGATCCGGAG GAGGAGGAGC
1251   CACAAACGAC GACGATGTTA AAAAAGCTGC CACTGTGGCC ATTGCTGCTG
1301   CCTACAACAA TGGCCAAGAA ATCAACGGTT TCAAAGCTGG AGAGACCATC
1351   TACGACATTG ATGAAGACGG CACAATTACC AAAAAAGACG CAACTGCAGC
1401   CGATGTTGAA GCCGACGACT TTAAAGGTCT GGGTCTGAAA AAAGTCGTGA
1451   CTAACCTGAC CAAAACCGTC AATGAAAACA AACAAAACGT CGATGCCAAA
1501   GTAAAAGCTG CAGAATCTGA AATAGAAAAG TTAACAACCA AGTTAGCAGA
1551   CACTGATGCC GCTTTAGCAG ATACTGATGC CGCTCTGGAT GCAACCACCA
1601   ACGCCTTGAA TAAATTGGGA GAAAATATAA CGACATTTGC TGAAGAGACT
1651   AAGACAAATA TCGTAAAAAT TGATGAAAAA TTAGAAGCCG TGGCTGATAC
1701   CGTCGACAAG CATGCCGAAG CATTCAACGA TATCGCCGAT TCATTGGATG
1751   AAACCAACAC TAAGGCAGAC GAAGCCGTCA AAACCGCCAA TGAAGCCAAA
1801   CAGACGGCCG AAGAAACCAA ACAAAACGTC GATGCCAAAG TAAAAGCTGC
1851   AGAAACTGCA GCAGGCAAAG CCGAAGCTGC CGCTGGCACA GCTAATACTG
```

```
1901  CAGCCGACAA GGCCGAAGCT GTCGCTGCAA AAGTTACCGA CATCAAAGCT
1951  GATATCGCTA CGAACAAAGA TAATATTGCT AAAAAAGCAA ACAGTGCCGA
2001  CGTGTACACC AGAGAAGAGT CTGACAGCAA ATTTGTCAGA ATTGATGGTC
2051  TGAACGCTAC TACCGAAAAA TTGGACACAC GCTTGGCTTC TGCTGAAAAA
2101  TCCATTGCCG ATCACGATAC TCGCCTGAAC GGTTTGGATA AAACAGTGTC
2151  AGACCTGCGC AAAGAAACCC GCCAAGGCCT TGCAGAACAA GCCGCGCTCT
2201  CCGGTCTGTT CCAACCTTAC AACGTGGGTC TCGAGCACCA CCACCACCAC
2251  CACTGA

   1  MSDLANDSFI RQVLDRQHFE PDGKYHLFGS RGELAERSGH IGLGKIQSHQ
  51  LGNLMIQQAA IKGNIGYIVR FSDHGHEVHS PFDNHASHSD SDEAGSPVDG
 101  FSLYRIHWDG YEHHPADGYD GPQGGGYPAP KGARDIYSYD IKGVAQNIRL
 151  NLTDNRSTGQ RLADRFHNAG SMLTQGVGDG FKRATRYSPE LDRSGNAAEA
 201  FNGTADIVKN IIGAAGEIVG AGDAVQGISE GSNIAVMHGL GLLSTENKMA
 251  RINDLADMAQ LKDYAAAAIR DWAVQNPNAA QGIEAVSNIF MAAIPIKGIG
 301  AVRGKYGLGG ITAHPIKRSQ MGAIALPKGK SAVSDNFADA AYAKYPSPYH
 351  SRNIRSNLEQ RYGKENITSS TVPPSNGKNV KLADQRHPKT GVPFDGKGFP
 401  NFEKHVKYDT GSGGGGATND DDVKKAATVA IAAAYNNGQE INGFKAGETI
 451  YDIDEDGTIT KKDATAADVE ADDFKGLGLK KVVTNLTKTV NENKQNVDAK
 501  VKAAESEIEK LTTKLADTDA ALADTDAALD ATTNALNKLG ENITTFAEET
 551  KTNIVKIDEK LEAVADTVDK HAEAFNDIAD SLDETNTKAD EAVKTANEAK
 601  QTAEETKQNV DAKVKAAETA AGKAEAAAGT ANTAADKAEA VAAKVTDIKA
 651  DIATNKDNIA KKANSADVYT REESDSKFVR IDGLNATTEK LDTRLASAEK
 701  SIADHDTRLN GLDKTVSDLR KETRQGLAEQ AALSGLFQPY NVGLEHHHHH
 751  H*
```

961-ORF46.1

```
   1  ATGGCCACAA ACGACGACGA TGTTAAAAAA GCTGCCACTG TGGCCATTGC
  51  TGCTGCCTAC AACAATGGCC AAGAAATCAA CGGTTTCAAA GCTGGAGAGA
 101  CCATCTACGA CATTGATGAA GACGGCACAA TTACCAAAAA AGACGCAACT
 151  GCAGCCGATG TTGAAGCCGA CGACTTTAAA GGTCTGGGTC TGAAAAAAGT
 201  CGTGACTAAC CTGACCAAAA CCGTCAATGA AAACAAACAA AACGTCGATG
 251  CCAAAGTAAA AGCTGCAGAA TCTGAAATAG AAAAGTTAAC AACCAAGTTA
 301  GCAGACACTG ATGCCGCTTT AGCAGATACT GATGCCGCTC TGGATGCAAC
 351  CACCAACGCC TTGAATAAAT TGGGAGAAAA TATAACGACA TTTGCTGAAG
 401  AGACTAAGAC AAATATCGTA AAAATTGATG AAAAATTAGA AGCCGTGGCT
 451  GATACCGTCG ACAAGCATGC CGAAGCATTC AACGATATCG CCGATTCATT
 501  GGATGAAACC AACACTAAGG CAGACGAAGC CGTCAAAACC GCCAATGAAG
 551  CCAAACAGAC GGCCGAAGAA ACCAAACAAA ACGTCGATGC CAAAGTAAAA
 601  GCTGCAGAAA CTGCAGCAGG CAAAGCCGAA GCTGCCGCTG GCACAGCTAA
 651  TACTGCAGCC GACAAGGCCG AAGCTGTCGC TGCAAAAGTT ACCGACATCA
 701  AAGCTGATAT CGCTACGAAC AAAGATAATA TTGCTAAAAA AGCAAACAGT
 751  GCCGACGTGT ACACCAGAGA AGAGTCTGAC AGCAAATTTG TCAGAATTGA
 801  TGGTCTGAAC GCTACTACCG AAAAATTGGA CACACGCTTG GCTTCTGCTG
 851  AAAAATCCAT TGCCGATCAC GATACTCGCC TGAACGGTTT GGATAAAACA
 901  GTGTCAGACC TGCGCAAAGA AACCCGCCAA GGCCTTGCAG AACAAGCCGC
 951  GCTCTCCGGT CTGTTCCAAC CTTACAACGT GGGTCGGTTC AATGTAACGG
1001  CTGCAGTCGG CGGCTACAAA TCCGAATCGG CAGTCGCCAT CGGTACCGGC
1051  TTCCGCTTTA CCGAAAACTT TGCCGCCAAA GCAGGCGTGG CAGTCGGCAC
1101  TTCGTCCGGT TCTTCCGCAG CCTACCATGT CGGCGTCAAT TACGAGTGGG
1151  GATCCGGAGG AGGAGGATCA GATTTGGCAA ACGATTCTTT TATCCGGCAG
1201  GTTCTCGACC GTCAGCATTT CGAACCCGAC GGGAAATACC ACCTATTCGG
1251  CAGCAGGGGG GAACTTGCCG AGCGCAGCGG CCATATCGGA TTGGGAAAAA
1301  TACAAAGCCA TCAGTTGGGC AACCTGATGA TTCAACAGGC GGCCATTAAA
1351  GGAAATATCG GCTACATTGT CCGCTTTTCC GATCACGGGC ACGAAGTCCA
1401  TTCCCCCTTC GACAACCATG CCTCACATTC CGATTCTGAT GAAGCCGGTA
1451  GTCCCGTTGA CGGATTTAGC CTTTACCGCA TCCATTGGGA CGGATACGAA
1501  CACCATCCCG CCGACGGCTA TGACGGGCCA CAGGGCGGCG GCTATCCCGC
1551  TCCCAAAGGC GCGAGGGATA TATACAGCTA CGACATAAAA GGCGTTGCCC
1601  AAAATATCCG CCTCAACCTG ACCGACAACC GCAGCACCGG ACAACGGCTT
1651  GCCGACCGTT TCCACAATGC CGGTAGTATG CTGACGCAAG GAGTAGGCGA
1701  CGGATTCAAA CGCGCCACCC GATACAGCCC CGAGCTGGAC AGATCGGGCA
1751  ATGCCGCCGA AGCCTTCAAC GGCACTGCAG ATATCGTTAA AAACATCATC
1801  GGCGCGGCAG GAGAAATTGT CGGCGCAGGC GATGCCGTGC AGGGCATAAG
1851  CGAAGGCTCA AACATTGCTG TCATGCACGG CTTGGGTCTG CTTTCCACCG
1901  AAAACAAGAT GGCGCGCATC AACGATTTGG CAGATATGGC GCAACTCAAA
```

```
1951   GACTATGCCG CAGCAGCCAT CCGCGATTGG GCAGTCCAAA ACCCCAATGC
2001   CGCACAAGGC ATAGAAGCCG TCAGCAATAT CTTTATGGCA GCCATCCCCA
2051   TCAAAGGGAT TGGAGCTGTT CGGGGAAAAT ACGGCTTGGG CGGCATCACG
2101   GCACATCCTA TCAAGCGGTC GCAGATGGGC GCGATCGCAT TGCCGAAAGG
2151   GAAATCCGCC GTCAGCGACA ATTTTGCCGA TGCGGCATAC GCCAAATACC
2201   CGTCCCCTTA CCATTCCCGA AATATCCGTT CAAACTTGGA GCAGCGTTAC
2251   GGCAAAGAAA ACATCACCTC CTCAACCGTG CCGCCGTCAA ACGGCAAAAA
2301   TGTCAAACTG GCAGACCAAC GCCACCCGAA GACAGGCGTA CCGTTTGACG
2351   GTAAAGGGTT TCCGAATTTT GAGAAGCACG TGAAATATGA TACGCTCGAG
2401   CACCACCACC ACCACCACTG A

1      MATNDDDVKK AATVAIAAAY NNGQEINGFK AGETIYDIDE DGTITKKDAT
51     AADVEADDFK GLGLKKVVTN LTKTVNENKQ NVDAKVKAAE SEIEKLTTKL
101    ADTDAALADT DAALDATTNA LNKLGENITT FAEETKTNIV KIDEKLEAVA
151    DTVDKHAEAF NDIADSLDET NTKADEAVKT ANEAKQTAEE TKQNVDAKVK
201    AAETAAGKAE AAAGTANTAA DKAEAVAAKV TDIKADIATN KDNIAKKANS
251    ADVYTREESD SKFVRIDGLN ATTEKLDTRL ASAEKSIADH DTRLNGLDKT
301    VSDLRKETRQ GLAEQAALSG LFQPYNVGRF NVTAAVGGYK SESAVAIGTG
351    FRFTENFAAK AGVAVGTSSG SSAAYHVGVN YEWGSGGGGS DLANDSFIRQ
401    VLDRQHFEPD GKYHLFGSRG ELAERSGHIG LGKIQSHQLG NLMIQQAAIK
451    GNIGYIVRFS DHGHEVHSPF DNHASHSDSD EAGSPVDGFS LYRIHWDGYE
501    HHPADGYDGP QGGGYPAPKG ARDIYSYDIK GVAQNIRLNL TDNRSTGQRL
551    ADRFHNAGSM LTQGVGDGFK RATRYSPELD RSGNAAEAFN GTADIVKNII
601    GAAGEIVGAG DAVQGISEGS NIAVMHGLGL LSTENKMARI NDLADMAQLK
651    DYAAAAIRDW AVQNPNAAQG IEAVSNIFMA AIPIKGIGAV RGKYGLGGIT
701    AHPIKRSQMG AIALPKGKSA VSDNFADAAY AKYPSPYHSR NIRSNLEQRY
751    GKENITSSTV PPSNGKNVKL ADQRHPKTGV PFDGKGFPNF EKHVKYDTLE
801    HHHHHH*

961-741
1      ATGGCCACAA ACGACGACGA TGTTAAAAAA GCTGCCACTG TGGCCATTGC
51     TGCTGCCTAC AACAATGGCC AAGAAATCAA CGGTTTCAAA GCTGGAGAGA
101    CCATCTACGA CATTGATGAA GACGGCACAA TTACCAAAAA AGACGCAACT
151    GCAGCCGATG TTGAAGCCGA CGACTTTAAA GGTCTGGGTC TGAAAAAAGT
201    CGTGACTAAC CTGACCAAAA CCGTCAATGA AAACAAACAA AACGTCGATG
251    CCAAAGTAAA AGCTGCAGAA TCTGAAATAG AAAAGTTAAC AACCAAGTTA
301    GCAGACACTG ATGCCGCTTT AGCAGATACT GATGCCGCTC TGGATGCAAC
351    CACCAACGCC TTGAATAAAT TGGGAGAAAA TATAACGACA TTTGCTGAAG
401    AGACTAAGAC AAATATCGTA AAAATTGATG AAAAATTAGA AGCCGTGGCT
451    GATACCGTCG ACAAGCATGC CGAAGCATTC AACGATATCG CCGATTCATT
501    GGATGAAACC AACACTAAGG CAGACGAAGC CGTCAAAACC GCCAATGAAG
551    CCAAACAGAC GGCCGAAGAA ACCAAACAAA ACGTCGATGC CAAAGTAAAA
601    GCTGCAGAAA CTGCAGCAGG CAAAGCCGAA GCTGCCGCTG GCACAGCTAA
651    TACTGCAGCC GACAAGGCCG AAGCTGTCGC TGCAAAAGTT ACCGACATCA
701    AAGCTGATAT CGCTACGAAC AAAGATAATA TTGCTAAAAA AGCAAACAGT
751    GCCGACGTGT ACACCAGAGA AGAGTCTGAC AGCAAATTTG TCAGAATTGA
801    TGGTCTGAAC GCTACTACCG AAAAATTGGA CACACGCTTG GCTTCTGCTG
851    AAAAATCCAT TGCCGATCAC GATACTCGCC TGAACGGTTT GGATAAAACA
901    GTGTCAGACC TGCGCAAAGA AACCCGCCAA GGCCTTGCAG AACAAGCCGC
951    GCTCTCCGGT CTGTTCCAAC CTTACAACGT GGGTCGGTTC AATGTAACGG
1001   CTGCAGTCGG CGGCTACAAA TCCGAATCGG CAGTCGCCAT CGGTACCGGC
1051   TTCCGCTTTA CCGAAAACTT TGCCGCCAAA GCAGGCGTGG CAGTCGGCAC
1101   TTCGTCCGGT TCTTCCGCAG CCTACCATGT CGGCGTCAAT TACGAGTGGG
1151   GATCCGGAGG GGGTGGTGTC GCCGCCGACA TCGGTGCGGG GCTTGCCGAT
1201   GCACTAACCG CACCGCTCGA CCATAAAGAC AAAGGTTTGC AGTCTTTGAC
1251   GCTGGATCAG TCCGTCAGGA AAAACGAGAA ACTGAAGCTG GCGGCACAAG
1301   GTGCGGAAAA AACTTATGGA AACGGTGACA GCCTCAATAC GGGCAAATTG
1351   AAGAACGACA AGGTCAGCCG TTTCGACTTT ATCCGCCAAA TCGAAGTGGA
1401   CGGGCAGCTC ATTACCTTGG AGAGTGGAGA GTTCCAAGTA TACAAACAAA
1451   GCCATTCCGC CTTAACCGCC TTTCAGACCG AGCAAATACA AGATTCGGAG
1501   CATTCCGGGA AGATGGTTGC GAAACGCCAG TTCAGAATCG GCGACATAGC
1551   GGGCGAACAT ACATCTTTTG ACAAGCTTCC CGAAGGCGGC AGGGCGACAT
1601   ATCGCGGGAC GGCGTTCGGT TCAGACGATG CCGGCGGAAA ACTGACCTAC
1651   ACCATAGATT TCGCCGCCAA GCAGGGAAAC GGCAAAATCG AACATTTGAA
1701   ATCGCCAGAA CTCAATGTCG ACCTGGCCGC CGCCGATATC AAGCCGGATG
1751   GAAAACGCCA TGCCGTCATC AGCGGTTCCG TCCTTTACAA CCAAGCCGAG
1801   AAAGGCAGTT ACTCCCTCGG TATCTTTGGC GGAAAAGCCC AGGAAGTTGC
```

```
1851  CGGCAGCGCG GAAGTGAAAA CCGTAAACGG CATACGCCAT ATCGGCCTTG
1901  CCGCCAAGCA ACTCGAGCAC CACCACCACC ACCACTGA

   1  MATNDDDVKK AATVAIAAAY NNGQEINGFK AGETIYDIDE DGTITKKDAT
  51  AADVEADDFK GLGLKKVVTN LTKTVNENKQ NVDAKVKAAE SEIEKLTTKL
 101  ADTDAALADT DAALDATTNA LNKLGENITT FAEETKTNIV KIDEKLEAVA
 151  DTVDKHAEAF NDIADSLDET NTKADEAVKT ANEAKQTAEE TKQNVDAKVK
 201  AAETAAGKAE AAAGTANTAA DKAEAVAAKV TDIKADIATN KDNIAKKANS
 251  ADVYTREESD SKFVRIDGLN ATTEKLDTRL ASAEKSIADH DTRLNGLDKT
 301  VSDLRKETRQ GLAEQAALSG LFQPYNVGRF NVTAAVGGYK SESAVAIGTG
 351  FRFTENFAAK AGVAVGTSSG SSAAYHVGVN YEWGSGGGGV AADIGAGLAD
 401  ALTAPLDHKD KGLQSLTLDQ SVRKNEKLKL AAQGAEKTYG NGDSLNTGKL
 451  KNDKVSRFDF IRQIEVDGQL ITLESGEFQV YKQSHSALTA FQTEQIQDSE
 501  HSGKMVAKRQ FRIGDIAGEH TSFDKLPEGG RATYRGTAFG SDDAGGKLTY
 551  TIDFAAKQGN GKIEHLKSPE LNVDLAAADI KPDGKRHAVI SGSVLYNQAE
 601  KGSYSLGIFG GKAQEVAGSA EVKTVNGIRH IGLAAKQLEH HHHHH*
```

961-983

```
   1  ATGGCCACAA ACGACGACGA TGTTAAAAAA GCTGCCACTG TGGCCATTGC
  51  TGCTGCCTAC AACAATGGCC AAGAAATCAA CGGTTTCAAA GCTGGAGAGA
 101  CCATCTACGA CATTGATGAA GACGGCACAA TTACCAAAAA AGACGCAACT
 151  GCAGCCGATG TTGAAGCCGA CGACTTTAAA GGTCTGGGTC TGAAAAAAGT
 201  CGTGACTAAC CTGACCAAAA CCGTCAATGA AAACAAACAA AACGTCGATG
 251  CCAAAGTAAA AGCTGCAGAA TCTGAAATAG AAAAGTTAAC AACCAAGTTA
 301  GCAGACACTG ATGCCGCTTT AGCAGATACT GATGCCGCTC TGGATGCAAC
 351  CACCAACGCC TTGAATAAAT TGGGAGAAAA TATAACGACA TTTGCTGAAG
 401  AGACTAAGAC AAATATCGTA AAAATTGATG AAAAATTAGA AGCCGTGGCT
 451  GATACCGTCG ACAAGCATGC CGAAGCATTC AACGATATCG CCGATTCATT
 501  GGATGAAACC AACACTAAGG CAGACGAAGC CGTCAAAACC GCCAATGAAG
 551  CCAAACAGAC GGCCGAAGAA ACCAAACAAA ACGTCGATGC CAAAGTAAAA
 601  GCTGCAGAAA CTGCAGCAGG CAAAGCCGAA GCTGCCGCTG GCACAGCTAA
 651  TACTGCAGCC GACAAGGCCG AAGCTGTCGC TGCAAAAGTT ACCGACATCA
 701  AAGCTGATAT CGCTACGAAC AAAGATAATA TTGCTAAAAA AGCAAACAGT
 751  GCCGACGTGT ACACCAGAGA AGAGTCTGAC AGCAAATTTG TCAGAATTGA
 801  TGGTCTGAAC GCTACTACCG AAAAATTGGA CACACGCTTG GCTTCTGCTG
 851  AAAAATCCAT TGCCGATCAC GATACTCGCC TGAACGGTTT GGATAAAACA
 901  GTGTCAGACC TGCGCAAAGA AACCCGCCAA GGCCTTGCAG AACAAGCCGC
 951  GCTCTCCGGT CTGTTCCAAC CTTACAACGT GGGTCGGTTC AATGTAACGG
1001  CTGCAGTCGG CGGCTACAAA TCCGAATCGG CAGTCGCCAT CGGTACCGGC
1051  TTCCGCTTTA CCGAAAACTT TGCCGCCAAA GCAGGCGTGG CAGTCGGCAC
1101  TTCGTCCGGT TCTTCCGCAG CCTACCATGT CGGCGTCAAT TACGAGTGGG
1151  GATCCGGCGG AGGCGGCACT TCTGCGCCCG ACTTCAATGC AGGCGGTACC
1201  GGTATCGGCA GCAACAGCAG AGCAACAACA GCGAAATCAG CAGCAGTATC
1251  TTACGCCGGT ATCAAGAACG AAATGTGCAA AGACAGAAGC ATGCTCTGTG
1301  CCGGTCGGGA TGACGTTGCG GTTACAGACA GGGATGCCAA AATCAATGCC
1351  CCCCCCCCGA ATCTGCATAC CGGAGACTTT CCAAACCCAA ATGACGCATA
1401  CAAGAATTTG ATCAACCTCA AACCTGCAAT TGAAGCAGGC TATACAGGAC
1451  GCGGGGTAGA GGTAGGTATC GTCGACACAG GCGAATCCGT CGGCAGCATA
1501  TCCTTTCCCG AACTGTATGG CAGAAAAGAA CACGGCTATA ACGAAAATTA
1551  CAAAAACTAT ACGGCGTATA TGCGGAAGGA AGCGCCTGAA GACGGAGGCG
1601  GTAAAGACAT TGAAGCTTCT TTCGACGATG AGGCCGTTAT AGAGACTGAA
1651  GCAAAGCCGA CGGATATCCG CCACGTAAAA GAAATCGGAC ACATCGATTT
1701  GGTCTCCCAT ATTATTGGCG GGCGTTCCGT GGACGGCAGA CCTGCAGGCG
1751  GTATTGCGCC CGATGCGACG CTACACATAA TGAATACGAA TGATGAAACC
1801  AAGAACGAAA TGATGGTTGC AGCCATCCGC AATGCATGGG TCAAGCTGGG
1851  CGAACGTGGC GTGCGCATCG TCAATAACAG TTTTGGAACA ACATCGAGGG
1901  CAGGCACTGC CGACCTTTTC CAAATAGCCA ATTCGGAGGA GCAGTACCGC
1951  CAAGCGTTGC TCGACTATTC CGGCGGTGAT AAAACAGACG AGGGTATCCG
2001  CCTGATGCAA CAGAGCGATT ACGGCAACCT GTCCTACCAC ATCCGTAATA
2051  AAAACATGCT TTTCATCTTT TCGACAGGCA ATGACGCACA AGCTCAGCCC
2101  AACACATATG CCCTATTGCC ATTTTATGAA AAAGACGCTC AAAAAGGCAT
2151  TATCACAGTC GCAGGCGTAG ACCGCAGTGG AGAAAAGTTC AAACGGGAAA
2201  TGTATGGAGA ACCGGGTACA GAACCGCTTG AGTATGGCTC CAACCATTGC
2251  GGAATTACTG CCATGTGGTG CCTGTCGGCA CCCTATGAAG CAAGCGTCCG
2301  TTTCACCCGT ACAAACCCGA TTCAAATTGC CGGAACATCC TTTTCCGCAC
2351  CCATCGTAAC CGGCACGGCG GCTCTGCTGC TGCAGAAATA CCCGTGGATG
```

```
2401  AGCAACGACA ACCTGCGTAC CACGTTGCTG ACGACGGCTC AGGACATCGG
2451  TGCAGTCGGC GTGGACAGCA AGTTCGGCTG GGGACTGCTG GATGCGGGTA
2501  AGGCCATGAA CGGACCCGCG TCCTTTCCGT TCGGCGACTT TACCGCCGAT
2551  ACGAAAGGTA CATCCGATAT TGCCTACTCC TTCCGTAACG ACATTTCAGG
2601  CACGGGCGGC CTGATCAAAA AAGGCGGCAG CCAACTGCAA CTGCACGGCA
2651  ACAACACCTA TACGGGCAAA ACCATTATCG AAGGCGGTTC GCTGGTGTTG
2701  TACGGCAACA ACAAATCGGA TATGCGCGTC GAAACCAAAG GTGCGCTGAT
2751  TTATAACGGG GCGGCATCCG GCGGCAGCCT GAACAGCGAC GGCATTGTCT
2801  ATCTGGCAGA TACCGACCAA TCCGGCGCAA ACGAAACCGT ACACATCAAA
2851  GGCAGTCTGC AGCTGGACGG CAAAGGTACG CTGTACACAC GTTTGGGCAA
2901  ACTGCTGAAA GTGGACGGTA CGGCGATTAT CGGCGGCAAG CTGTACATGT
2951  CGGCACGCGG CAAGGGGGCA GGCTATCTCA ACAGTACCGG ACGACGTGTT
3001  CCCTTCCTGA GTGCCGCCAA AATCGGGCAG GATTATTCTT TCTTCACAAA
3051  CATCGAAACC GACGGCGGCC TGCTGGCTTC CCTCGACAGC GTCGAAAAAA
3101  CAGCGGGCAG TGAAGGCGAC ACGCTGTCCT ATTATGTCCG TCGCGGCAAT
3151  GCGGCACGGA CTGCTTCGGC AGCGGCACAT TCCGCGCCCG CCGGTCTGAA
3201  ACACGCCGTA GAACAGGGCG GCAGCAATCT GGAAAACCTG ATGGTCGAAC
3251  TGGATGCCTC CGAATCATCC GCAACACCCG AGACGGTTGA AACTGCGGCA
3301  GCCGACCGCA CAGATATGCC GGGCATCCGC CCCTACGGCG CAACTTTCCG
3351  CGCAGCGGCA GCCGTACAGC ATGCGAATGC CGCCGACGGT GTACGCATCT
3401  TCAACAGTCT CGCCGCTACC GTCTATGCCG ACAGTACCGC CGCCCATGCC
3451  GATATGCAGG GACGCCGCCT GAAAGCCGTA TCGGACGGGT TGGACCACAA
3501  CGGCACGGGT CTGCGCGTCA TCGCGCAAAC CCAACAGGAC GGTGGAACGT
3551  GGGAACAGGG CGGTGTTGAA GGCAAAATGC GCGGCAGTAC CCAAACCGTC
3601  GGCATTGCCG CGAAAACCGG CGAAAATACG ACAGCAGCCG CCACACTGGG
3651  CATGGGACGC AGCACATGGA GCGAAAACAG TGCAAATGCA AAAACCGACA
3701  GCATTAGTCT GTTTGCAGGC ATACGGCACG ATGCGGGCGA TATCGGCTAT
3751  CTCAAAGGCC TGTTCTCCTA CGGACGCTAC AAAAACAGCA TCAGCCGCAG
3801  CACCGGTGCG GACGAACATG CGGAAGGCAG CGTCAACGGC ACGCTGATGC
3851  AGCTGGGCGC ACTGGGCGGT GTCAACGTTC CGTTTGCCGC AACGGGAGAT
3901  TTGACGGTCG AAGGCGGTCT GCGCTACGAC CTGCTCAAAC AGGATGCATT
3951  CGCCGAAAAA GGCAGTGCTT TGGGCTGGAG CGGCAACAGC CTCACTGAAG
4001  GCACGCTGGT CGGACTCGCG GGTCTGAAGC TGTCGCAACC CTTGAGCGAT
4051  AAAGCCGTCC TGTTTGCAAC GGCGGGCGTG GAACGCGACC TGAACGGACG
4101  CGACTACACG GTAACGGGCG GCTTTACCGG CGCGACTGCA GCAACCGGCA
4151  AGACGGGGGC ACGCAATATG CCGCACACCC GTCTGGTTGC CGGCCTGGGC
4201  GCGGATGTCG AATTCGGCAA CGGCTGGAAC GGCTTGGCAC GTTACAGCTA
4251  CGCCGGTTCC AAACAGTACG GCAACCACAG CGGACGAGTC GGCGTAGGCT
4301  ACCGGTTCCT CGAGCACCAC CACCACCACC ACTGA

   1  MATNDDDVKK AATVAIAAAY NNGQEINGFK AGETIYDIDE DGTITKKDAT
  51  AADVEADDFK GLGLKKVVTN LTKTVNENKQ NVDAKVKAAE SEIEKLTTKL
 101  ADTDAALADT DAALDATTNA LNKLGENITT FAEETKTNIV KIDEKLEAVA
 151  DTVDKHAEAF NDIADSLDET NTKADEAVKT ANEAKQTAEE TKQNVDAKVK
 201  AAETAAGKAE AAAGTANTAA DKAEAVAAKV TDIKADIATN KDNIAKKANS
 251  ADVYTREESD SKFVRIDGLN ATTEKLDTRL ASAEKSIADH DTRLNGLDKT
 301  VSDLRKETRQ GLAEQAALSG LFQPYNVGRF NVTAAVGGYK SESAVAIGTG
 351  FRFTENFAAK AGVAVGTSSG SSAAYHVGVN YEWGSGGGGT SAPDFNAGGT
 401  GIGSNSRATT AKSAAVSYAG IKNEMCKDRS MLCAGRDDVA VTDRDAKINA
 451  PPPNLHTGDF PNPNDAYKNL INLKPAIEAG YTGRGVEVGI VDTGESVGSI
 501  SFPELYGRKE HGYNENYKNY TAYMRKEAPE DGGGKDIEAS FDDEAVIETE
 551  AKPTDIRHVK EIGHIDLVSH IIGGRSVDGR PAGGIAPDAT LHIMNTNDET
 601  KNEMMVAAIR NAWVKLGERG VRIVNNSFGT TSRAGTADLF QIANSEEQYR
 651  QALLDYSGGD KTDEGIRLMQ QSDYGNLSYH IRNKNMLFIF STGNDAQAQP
 701  NTYALLPFYE KDAQKGIITV AGVDRSGEKF KREMYGEPGT EPLEYGSNHC
 751  GITAMWCLSA PYEASVRFTR TNPIQIAGTS FSAPIVTGTA ALLLQKYPWM
 801  SNDNLRTTLL TTAQDIGAVG VDSKFGWGLL DAGKAMNGPA SFPFGDFTAD
 851  TKGTSDIAYS FRNDISGTGG LIKKGGSQLQ LHGNNTYTGK TIIEGGSLVL
 901  YGNNKSDMRV ETKGALIYNG AASGGSLNSD GIVYLADTDQ SGANETVHIK
 951  GSLQLDGKGT LYTRLGKLLK VDGTAIIGGK LYMSARGKGA GYLNSTGRRV
1001  PFLSAAKIGQ DYSFFTNIET DGGLLASLDS VEKTAGSEGD TLSYYVRRGN
1051  AARTASAAAH SAPAGLKHAV EQGGSNLENL MVELDASESS ATPETVETAA
1101  ADRTDMPGIR PYGATFRAAA AVQHANAADG VRIFNSLAAT VYADSTAAHA
1151  DMQGRRLKAV SDGLDHNGTG LRVIAQTQQD GGTWEQGGVE GKMRGSTQTV
1201  GIAAKTGENT TAAATLGMGR STWSENSANA KTDSISLFAG IRHDAGDIGY
1251  LKGLFSYGRY KNSISRSTGA DEHAEGSVNG TLMQLGALGG VNVPFAATGD
1301  LTVEGGLRYD LLKQDAFAEK GSALGWSGNS LTEGTLVGLA GLKLSQPLSD
```

```
1351   KAVLFATAGV ERDLNGRDYT VTGGFTGATA ATGKTGARNM PHTRLVAGLG
1401   ADVEFGNGWN GLARYSYAGS KQYGNHSGRV GVGYRFLEHH HHHH*
```

961c-ORF46.1

```
   1   ATGGCCACAA ACGACGACGA TGTTAAAAAA GCTGCCACTG TGGCCATTGC
  51   TGCTGCCTAC AACAATGGCC AAGAAATCAA CGGTTTCAAA GCTGGAGAGA
 101   CCATCTACGA CATTGATGAA GACGGCACAA TTACCAAAAA AGACGCAACT
 151   GCAGCCGATG TTGAAGCCGA CGACTTTAAA GGTCTGGGTC TGAAAAAAGT
 201   CGTGACTAAC CTGACCAAAA CCGTCAATGA AAACAAACAA AACGTCGATG
 251   CCAAAGTAAA AGCTGCAGAA TCTGAAATAG AAAAGTTAAC AACCAAGTTA
 301   GCAGACACTG ATGCCGCTTT AGCAGATACT GATGCCGCTC TGGATGCAAC
 351   CACCAACGCC TTGAATAAAT TGGGAGAAAA TATAACGACA TTTGCTGAAG
 401   AGACTAAGAC AAATATCGTA AAAATTGATG AAAAATTAGA AGCCGTGGCT
 451   GATACCGTCG ACAAGCATGC CGAAGCATTC AACGATATCG CCGATTCATT
 501   GGATGAAACC AACACTAAGG CAGACGAAGC CGTCAAAACC GCCAATGAAG
 551   CCAAACAGAC GGCCGAAGAA ACCAAACAAA ACGTCGATGC CAAAGTAAAA
 601   GCTGCAGAAA CTGCAGCAGG CAAAGCCGAA GCTGCCGCTG GCACAGCTAA
 651   TACTGCAGCC GACAAGGCCG AAGCTGTCGC TGCAAAAGTT ACCGACATCA
 701   AAGCTGATAT CGCTACGAAC AAAGATAATA TTGCTAAAAA AGCAAACAGT
 751   GCCGACGTGT ACACCAGAGA AGAGTCTGAC AGCAAATTTG TCAGAATTGA
 801   TGGTCTGAAC GCTACTACCG AAAAATTGGA CACACGCTTG GCTTCTGCTG
 851   AAAAATCCAT TGCCGATCAC GATACTCGCC TGAACGGTTT GGATAAAACA
 901   GTGTCAGACC TGCGCAAAGA AACCCGCCAA GGCCTTGCAG AACAAGCCGC
 951   GCTCTCCGGT CTGTTCCAAC CTTACAACGT GGGTGGATCC GGAGGAGGAG
1001   GATCAGATTT GGCAAACGAT TCTTTTATCC GGCAGGTTCT CGACCGTCAG
1051   CATTTCGAAC CCGACGGGAA ATACCACCTA TTCGGCAGCA GGGGGGAACT
1101   TGCCGAGCGC AGCGGCCATA TCGGATTGGG AAAAATACAA AGCCATCAGT
1151   TGGGCAACCT GATGATTCAA CAGGCGGCCA TTAAAGGAAA TATCGGCTAC
1201   ATTGTCCGCT TTTCCGATCA CGGGCACGAA GTCCATTCCC CCTTCGACAA
1251   CCATGCCTCA CATTCCGATT CTGATGAAGC CGGTAGTCCC GTTGACGGAT
1301   TTAGCCTTTA CCGCATCCAT TGGGACGGAT ACGAACACCA TCCCGCCGAC
1351   GGCTATGACG GGCCACAGGG CGGCGGCTAT CCCGCTCCCA AAGGCGCGAG
1401   GGATATATAC AGCTACGACA TAAAAGGCGT TGCCCAAAAT ATCCGCCTCA
1451   ACCTGACCGA CAACCGCAGC ACCGGACAAC GGCTTGCCGA CCGTTTCCAC
1501   AATGCCGGTA GTATGCTGAC GCAAGGAGTA GGCGACGGAT TCAAACGCGC
1551   CACCCGATAC AGCCCCGAGC TGGACAGATC GGGCAATGCC GCCGAAGCCT
1601   TCAACGGCAC TGCAGATATC GTTAAAAACA TCATCGGCGC GGCAGGAGAA
1651   ATTGTCGGCG CAGGCGATGC CGTGCAGGGC ATAAGCGAAG GCTCAAACAT
1701   TGCTGTCATG CACGGCTTGG GTCTGCTTTC CACCGAAAAC AAGATGGCGC
1751   GCATCAACGA TTTGGCAGAT ATGGCGCAAC TCAAAGACTA TGCCGCAGCA
1801   GCCATCCGCG ATTGGGCAGT CCAAAACCCC AATGCCGCAC AAGGCATAGA
1851   AGCCGTCAGC AATATCTTTA TGGCAGCCAT CCCCATCAAA GGGATTGGAG
1901   CTGTTCGGGG AAAATACGGC TTGGGCGGCA TCACGGCACA TCCTATCAAG
1951   CGGTCGCAGA TGGGCGCGAT CGCATTGCCG AAAGGGAAAT CCGCCGTCAG
2001   CGACAATTTT GCCGATGCGG CATACGCCAA ATACCCGTCC CCTTACCATT
2051   CCCGAAATAT CCGTTCAAAC TTGGAGCAGC GTTACGGCAA AGAAAACATC
2101   ACCTCCTCAA CCGTGCCGCC GTCAAACGGC AAAAATGTCA AACTGGCAGA
2151   CCAACGCCAC CCGAAGACAG GCGTACCGTT TGACGGTAAA GGGTTTCCGA
2201   ATTTTGAGAA GCACGTGAAA TATGATACGC TCGAGCACCA CCACCACCAC
2251   CACTGA
```

```
   1   MATNDDDVKK AATVAIAAAY NNGQEINGFK AGETIYDIDE DGTITKKDAT
  51   AADVEADDFK GLGLKKVVTN LTKTVNENKQ NVDAKVKAAE SEIEKLTTKL
 101   ADTDAALADT DAALDATTNA LNKLGENITT FAEETKTNIV KIDEKLEAVA
 151   DTVDKHAEAF NDIADSLDET NTKADEAVKT ANEAKQTAEE TKQNVDAKVK
 201   AAETAAGKAE AAAGTANTAA DKAEAVAAKV TDIKADIATN KDNIAKKANS
 251   ADVYTREESD SKFVRIDGLN ATTEKLDTRL ASAEKSIADH DTRLNGLDKT
 301   VSDLRKETRQ GLAEQAALSG LFQPYNVGGS GGGGSDLAND SFIRQVLDRQ
 351   HFEPDGKYHL FGSRGELAER SGHIGLGKIQ SHQLGNLMIQ QAAIKGNIGY
 401   IVRFSDHGHE VHSPFDNHAS HSDSDEAGSP VDGFSLYRIH WDGYEHHPAD
 451   GYDGPQGGGY PAPKGARDIY SYDIKGVAQN IRLNLTDNRS TGQRLADRFH
 501   NAGSMLTQGV GDGFKRATRY SPELDRSGNA AEAFNGTADI VKNIIGAAGE
 551   IVGAGDAVQG ISEGSNIAVM HGLGLLSTEN KMARINDLAD MAQLKDYAAA
 601   AIRDWAVQNP NAAQGIEAVS NIFMAAIPIK GIGAVRGKYG LGGITAHPIK
 651   RSQMGAIALP KGKSAVSDNF ADAAYAKYPS PYHSRNIRSN LEQRYGKENI
 701   TSSTVPPSNG KNVKLADQRH PKTGVPFDGK GFPNFEKHVK YDTLEHHHHH
```

751 H*

961c-741

1 ATGGCCACAA ACGACGACGA TGTTAAAAAA GCTGCCACTG TGGCCATTGC
51 TGCTGCCTAC AACAATGGCC AAGAAATCAA CGGTTTCAAA GCTGGAGAGA
101 CCATCTACGA CATTGATGAA GACGGCACAA TTACCAAAAA AGACGCAACT
151 GCAGCCGATG TTGAAGCCGA CGACTTTAAA GGTCTGGGTC TGAAAAAAGT
201 CGTGACTAAC CTGACCAAAA CCGTCAATGA AAACAAACAA AACGTCGATG
251 CCAAAGTAAA AGCTGCAGAA TCTGAAATAG AAAAGTTAAC AACCAAGTTA
301 GCAGACACTG ATGCCGCTTT AGCAGATACT GATGCCGCTC TGGATGCAAC
351 CACCAACGCC TTGAATAAAT TGGGAGAAAA TATAACGACA TTTGCTGAAG
401 AGACTAAGAC AAATATCGTA AAAATTGATG AAAAATTAGA AGCCGTGGCT
451 GATACCGTCG ACAAGCATGC CGAAGCATTC AACGATATCG CCGATTCATT
501 GGATGAAACC AACACTAAGG CAGACGAAGC CGTCAAAACC GCCAATGAAG
551 CCAAACAGAC GGCCGAAGAA ACCAAACAAA ACGTCGATGC CAAAGTAAAA
601 GCTGCAGAAA CTGCAGCAGG CAAAGCCGAA GCTGCCGCTG GCACAGCTAA
651 TACTGCAGCC GACAAGGCCG AAGCTGTCGC TGCAAAAGTT ACCGACATCA
701 AAGCTGATAT CGCTACGAAC AAAGATAATA TTGCTAAAAA AGCAAACAGT
751 GCCGACGTGT ACACCAGAGA AGAGTCTGAC AGCAAATTTG TCAGAATTGA
801 TGGTCTGAAC GCTACTACCG AAAAATTGGA CACACGCTTG GCTTCTGCTG
851 AAAAATCCAT TGCCGATCAC GATACTCGCC TGAACGGTTT GGATAAAACA
901 GTGTCAGACC TGCGCAAAGA AACCCGCCAA GGCCTTGCAG AACAAGCCGC
951 GCTCTCCGGT CTGTTCCAAC CTTACAACGT GGGTGGATCC GGAGGGGGTG
1001 GTGTCGCCGC CGACATCGGT GCGGGGCTTG CCGATGCACT AACCGCACCG
1051 CTCGACCATA AAGACAAAGG TTTGCAGTCT TTGACGCTGG ATCAGTCCGT
1101 CAGGAAAAAC GAGAAACTGA AGCTGGCGGC ACAAGGTGCG GAAAAAACTT
1151 ATGGAAACGG TGACAGCCTC AATACGGGCA AATTGAAGAA CGACAAGGTC
1201 AGCCGTTTCG ACTTTATCCG CCAAATCGAA GTGGACGGGC AGCTCATTAC
1251 CTTGGAGAGT GGAGAGTTCC AAGTATACAA ACAAAGCCAT TCCGCCTTAA
1301 CCGCCTTTCA GACCGAGCAA ATACAAGATT CGGAGCATTC CGGGAAGATG
1351 GTTGCGAAAC GCCAGTTCAG AATCGGCGAC ATAGCGGGCG AACATACATC
1401 TTTTGACAAG CTTCCCGAAG GCGGCAGGGC GACATATCGC GGGACGGCGT
1451 TCGGTTCAGA CGATGCCGGC GGAAAACTGA CCTACACCAT AGATTTCGCC
1501 GCCAAGCAGG GAAACGGCAA AATCGAACAT TTGAAATCGC CAGAACTCAA
1551 TGTCGACCTG GCCGCCGCCG ATATCAAGCC GGATGGAAAA CGCCATGCCG
1601 TCATCAGCGG TTCCGTCCTT TACAACCAAG CCGAGAAAGG CAGTTACTCC
1651 CTCGGTATCT TTGGCGGAAA AGCCCAGGAA GTTGCCGGCA GCGCGGAAGT
1701 GAAAACCGTA AACGGCATAC GCCATATCGG CCTTGCCGCC AAGCAACTCG
1751 AGCACCACCA CCACCACCAC TGA

1 MATNDDDVKK AATVAIAAAY NNGQEINGFK AGETIYDIDE DGTITKKDAT
51 AADVEADDFK GLGLKKVVTN LTKTVNENKQ NVDAKVKAAE SEIEKLTTKL
101 ADTDAALADT DAALDATTNA LNKLGENITT FAEETKTNIV KIDEKLEAVA
151 DTVDKHAEAF NDIADSLDET NTKADEAVKT ANEAKQTAEE TKQNVDAKVK
201 AAETAAGKAE AAAGTANTAA DKAEAVAAKV TDIKADIATN KDNIAKKANS
251 ADVYTREESD SKFVRIDGLN ATTEKLDTRL ASAEKSIADH DTRLNGLDKT
301 VSDLRKETRQ GLAEQAALSG LFQPYNVGGS GGGGVAADIG AGLADALTAP
351 LDHKDKGLQS LTLDQSVRKN EKLKLAAQGA EKTYGNGDSL NTGKLKNDKV
401 SRFDFIRQIE VDGQLITLES GEFQVYKQSH SALTAFQTEQ IQDSEHSGKM
451 VAKRQFRIGD IAGEHTSFDK LPEGGRATYR GTAFGSDDAG GKLTYTIDFA
501 AKQGNGKIEH LKSPELNVDL AAADIKPDGK RHAVISGSVL YNQAEKGSYS
551 LGIFGGKAQE VAGSAEVKTV NGIRHIGLAA KQLEHHHHHH *

961c-983

1 ATGGCCACAA ACGACGACGA TGTTAAAAAA GCTGCCACTG TGGCCATTGC
51 TGCTGCCTAC AACAATGGCC AAGAAATCAA CGGTTTCAAA GCTGGAGAGA
101 CCATCTACGA CATTGATGAA GACGGCACAA TTACCAAAAA AGACGCAACT
151 GCAGCCGATG TTGAAGCCGA CGACTTTAAA GGTCTGGGTC TGAAAAAAGT
201 CGTGACTAAC CTGACCAAAA CCGTCAATGA AAACAAACAA AACGTCGATG
251 CCAAAGTAAA AGCTGCAGAA TCTGAAATAG AAAAGTTAAC AACCAAGTTA
301 GCAGACACTG ATGCCGCTTT AGCAGATACT GATGCCGCTC TGGATGCAAC
351 CACCAACGCC TTGAATAAAT TGGGAGAAAA TATAACGACA TTTGCTGAAG
401 AGACTAAGAC AAATATCGTA AAAATTGATG AAAAATTAGA AGCCGTGGCT
451 GATACCGTCG ACAAGCATGC CGAAGCATTC AACGATATCG CCGATTCATT
501 GGATGAAACC AACACTAAGG CAGACGAAGC CGTCAAAACC GCCAATGAAG

```
 551  CCAAACAGAC GGCCGAAGAA ACCAAACAAA ACGTCGATGC CAAAGTAAAA
 601  GCTGCAGAAA CTGCAGCAGG CAAAGCCGAA GCTGCCGCTG GCACAGCTAA
 651  TACTGCAGCC GACAAGGCCG AAGCTGTCGC TGCAAAAGTT ACCGACATCA
 701  AAGCTGATAT CGCTACGAAC AAAGATAATA TTGCTAAAAA AGCAAACAGT
 751  GCCGACGTGT ACACCAGAGA AGAGTCTGAC AGCAAATTTG TCAGAATTGA
 801  TGGTCTGAAC GCTACTACCG AAAAATTGGA CACACGCTTG GCTTCTGCTG
 851  AAAAATCCAT TGCCGATCAC GATACTCGCC TGAACGGTTT GGATAAAACA
 901  GTGTCAGACC TGCGCAAAGA AACCCGCCAA GGCCTTGCAG AACAAGCCGC
 951  GCTCTCCGGT CTGTTCCAAC CTTACAACGT GGGTGGATCC GGCGGAGGCG
1001  GCACTTCTGC GCCCGACTTC AATGCAGGCG GTACCGGTAT CGGCAGCAAC
1051  AGCAGAGCAA CAACAGCGAA ATCAGCAGCA GTATCTTACG CCGGTATCAA
1101  GAACGAAATG TGCAAAGACA GAAGCATGCT CTGTGCCGGT CGGGATGACG
1151  TTGCGGTTAC AGACAGGGAT GCCAAAATCA ATGCCCCCCC CCCGAATCTG
1201  CATACCGGAG ACTTTCCAAA CCCAAATGAC GCATACAAGA ATTTGATCAA
1251  CCTCAAACCT GCAATTGAAG CAGGCTATAC AGGACGCGGG GTAGAGGTAG
1301  GTATCGTCGA CACAGGCGAA TCCGTCGGCA GCATATCCTT TCCCGAACTG
1351  TATGGCAGAA AAGAACACGG CTATAACGAA AATTACAAAA ACTATACGGC
1401  GTATATGCGG AAGGAAGCGC CTGAAGACGG AGGCGGTAAA GACATTGAAG
1451  CTTCTTTCGA CGATGAGGCC GTTATAGAGA CTGAAGCAAA GCCGACGGAT
1501  ATCCGCCACG TAAAAGAAAT CGGACACATC GATTTGGTCT CCCATATTAT
1551  TGGCGGGCGT TCCGTGGACG GCAGACCTGC AGGCGGTATT GCGCCCGATG
1601  CGACGCTACA CATAATGAAT ACGAATGATG AAACCAAGAA CGAAATGATG
1651  GTTGCAGCCA TCCGCAATGC ATGGGTCAAG CTGGGCGAAC GTGGCGTGCG
1701  CATCGTCAAT AACAGTTTTG GAACAACATC GAGGGCAGGC ACTGCCGACC
1751  TTTTCCAAAT AGCCAATTCG GAGGAGCAGT ACCGCCAAGC GTTGCTCGAC
1801  TATTCCGGCG GTGATAAAAC AGACGAGGGT ATCCGCCTGA TGCAACAGAG
1851  CGATTACGGC AACCTGTCCT ACCACATCCG TAATAAAAAC ATGCTTTTCA
1901  TCTTTTCGAC AGGCAATGAC GCACAAGCTC AGCCCAACAC ATATGCCCTA
1951  TTGCCATTTT ATGAAAAAGA CGCTCAAAAA GGCATTATCA CAGTCGCAGG
2001  CGTAGACCGC AGTGGAGAAA AGTTCAAACG GGAAATGTAT GGAGAACCGG
2051  GTACAGAACC GCTTGAGTAT GGCTCCAACC ATTGCGGAAT TACTGCCATG
2101  TGGTGCCTGT CGGCACCCTA TGAAGCAAGC GTCCGTTTCA CCCGTACAAA
2151  CCCGATTCAA ATTGCCGGAA CATCCTTTTC CGCACCCATC GTAACCGGCA
2201  CGGCGGCTCT GCTGCTGCAG AAATACCCGT GGATGAGCAA CGACAACCTG
2251  CGTACCACGT TGCTGACGAC GGCTCAGGAC ATCGGTGCAG TCGGCGTGGA
2301  CAGCAAGTTC GGCTGGGGAC TGCTGGATGC GGGTAAGGCC ATGAACGGAC
2351  CCGCGTCCTT TCCGTTCGGC GACTTTACCG CCGATACGAA AGGTACATCC
2401  GATATTGCCT ACTCCTTCCG TAACGACATT TCAGGCACGG GCGGCCTGAT
2451  CAAAAAAGGC GGCAGCCAAC TGCAACTGCA CGGCAACAAC ACCTATACGG
2501  GCAAAACCAT TATCGAAGGC GGTTCGCTGG TGTTGTACGG CAACAACAAA
2551  TCGGATATGC GCGTCGAAAC CAAAGGTGCG CTGATTTATA ACGGGGCGGC
2601  ATCCGGCGGC AGCCTGAACA GCGACGGCAT TGTCTATCTG GCAGATACCG
2651  ACCAATCCGG CGCAAACGAA ACCGTACACA TCAAAGGCAG TCTGCAGCTG
2701  GACGGCAAAG GTACGCTGTA CACACGTTTG GGCAAACTGC TGAAAGTGGA
2751  CGGTACGGCG ATTATCGGCG GCAAGCTGTA CATGTCGGCA CGCGGCAAGG
2801  GGGCAGGCTA TCTCAACAGT ACCGGACGAC GTGTTCCCTT CCTGAGTGCC
2851  GCCAAAATCG GGCAGGATTA TTCTTTCTTC ACAAACATCG AAACCGACGG
2901  CGGCCTGCTG GCTTCCCTCG ACAGCGTCGA AAAAACAGCG GGCAGTGAAG
2951  GCGACACGCT GTCCTATTAT GTCCGTCGCG GCAATGCGGC ACGGACTGCT
3001  TCGGCAGCGG CACATTCCGC GCCCGCCGGT CTGAAACACG CCGTAGAACA
3051  GGGCGGCAGC AATCTGGAAA ACCTGATGGT CGAACTGGAT GCCTCCGAAT
3101  CATCCGCAAC ACCCGAGACG GTTGAAACTG CGGCAGCCGA CCGCACAGAT
3151  ATGCCGGGCA TCCGCCCCTA CGGCGCAACT TTCCGCGCAG CGGCAGCCGT
3201  ACAGCATGCG AATGCCGCCG ACGGTGTACG CATCTTCAAC AGTCTCGCCG
3251  CTACCGTCTA TGCCGACAGT ACCGCCGCCC ATGCCGATAT GCAGGGACGC
3301  CGCCTGAAAG CCGTATCGGA CGGGTTGGAC CACAACGGCA CGGGTCTGCG
3351  CGTCATCGCG CAAACCCAAC AGGACGGTGG AACGTGGGAA CAGGGCGGTG
3401  TTGAAGGCAA AATGCGCGGC AGTACCCAAA CCGTCGGCAT TGCCGCGAAA
3451  ACCGGCGAAA ATACGACAGC AGCCGCCACA CTGGGCATGG GACGCAGCAC
3501  ATGGAGCGAA AACAGTGCAA ATGCAAAAAC CGACAGCATT AGTCTGTTTG
3551  CAGGCATACG GCACGATGCG GGCGATATCG GCTATCTCAA AGGCCTGTTC
3601  TCCTACGGAC GCTACAAAAA CAGCATCAGC CGCAGCACCG GTGCGGACGA
3651  ACATGCGGAA GGCAGCGTCA ACGGCACGCT GATGCAGCTG GGCGCACTGG
3701  GCGGTGTCAA CGTTCCGTTT GCCGCAACGG GAGATTTGAC GGTCGAAGGC
3751  GGTCTGCGCT ACGACCTGCT CAAACAGGAT GCATTCGCCG AAAAAGGCAG
3801  TGCTTTGGGC TGGAGCGGCA ACAGCCTCAC TGAAGGCACG CTGGTCGGAC
3851  TCGCGGGTCT GAAGCTGTCG CAACCCTTGA GCGATAAAGC CGTCCTGTTT
```

```
3901  GCAACGGCGG GCGTGGAACG CGACCTGAAC GGACGCGACT ACACGGTAAC
3951  GGGCGGCTTT ACCGGCGCGA CTGCAGCAAC CGGCAAGACG GGGGCACGCA
4001  ATATGCCGCA CACCCGTCTG GTTGCCGGCC TGGGCGCGGA TGTCGAATTC
4051  GGCAACGGCT GGAACGGCTT GGCACGTTAC AGCTACGCCG GTTCCAAACA
4101  GTACGGCAAC CACAGCGGAC GAGTCGGCGT AGGCTACCGG TTCCTCGAGC
4151  ACCACCACCA CCACCACTGA

   1  MATNDDDVKK AATVAIAAAY NNGQEINGFK AGETIYDIDE DGTITKKDAT
  51  AADVEADDFK GLGLKKVVTN LTKTVNENKQ NVDAKVKAAE SEIEKLTTKL
 101  ADTDAALADT DAALDATTNA LNKLGENITT FAEETKTNIV KIDEKLEAVA
 151  DTVDKHAEAF NDIADSLDET NTKADEAVKT ANEAKQTAEE TKQNVDAKVK
 201  AAETAAGKAE AAAGTANTAA DKAEAVAAKV TDIKADIATN KDNIAKKANS
 251  ADVYTREESD SKFVRIDGLN ATTEKLDTRL ASAEKSIADH DTRLNGLDKT
 301  VSDLRKETRQ GLAEQAALSG LFQPYNVGGS GGGGTSAPDF NAGGTGIGSN
 351  SRATTAKSAA VSYAGIKNEM CKDRSMLCAG RDDVAVTDRD AKINAPPPNL
 401  HTGDFPNPND AYKNLINLKP AIEAGYTGRG VEVGIVDTGE SVGSISFPEL
 451  YGRKEHGYNE NYKNYTAYMR KEAPEDGGGK DIEASFDDEA VIETEAKPTD
 501  IRHVKEIGHI DLVSHIIGGR SVDGRPAGGI APDATLHIMN TNDETKNEMM
 551  VAAIRNAWVK LGERGVRIVN NSFGTTSRAG TADLFQIANS EEQYRQALLD
 601  YSGGDKTDEG IRLMQQSDYG NLSYHIRNKN MLFIFSTGND AQAQPNTYAL
 651  LPFYEKDAQK GIITVAGVDR SGEKFKREMY GEPGTEPLEY GSNHCGITAM
 701  WCLSAPYEAS VRFTRTNPIQ IAGTSFSAPI VTGTAALLLQ KYPWMSNDNL
 751  RTTLLTTAQD IGAVGVDSKF GWGLLDAGKA MNGPASFPFG DFTADTKGTS
 801  DIAYSFRNDI SGTGGLIKKG GSQLQLHGNN TYTGKTIIEG GSLVLYGNNK
 851  SDMRVETKGA LIYNGAASGG SLNSDGIVYL ADTDQSGANE TVHIKGSLQL
 901  DGKGTLYTRL GKLLKVDGTA IIGGKLYMSA RGKGAGYLNS TGRRVPFLSA
 951  AKIGQDYSFF TNIETDGGLL ASLDSVEKTA GSEGDTLSYY VRRGNAARTA
1001  SAAAHSAPAG LKHAVEQGGS NLENLMVELD ASESSATPET VETAAADRTD
1051  MPGIRPYGAT FRAAAAVQHA NAADGVRIFN SLAATVYADS TAAHADMQGR
1101  RLKAVSDGLD HNGTGLRVIA QTQQDGGTWE QGGVEGKMRG STQTVGIAAK
1151  TGENTTAAAT LGMGRSTWSE NSANAKTDSI SLFAGIRHDA GDIGYLKGLF
1201  SYGRYKNSIS RSTGADEHAE GSVNGTLMQL GALGGVNVPF AATGDLTVEG
1251  GLRYDLLKQD AFAEKGSALG WSGNSLTEGT LVGLAGLKLS QPLSDKAVLF
1301  ATAGVERDLN GRDYTVTGGF TGATAATGKT GARNMPHTRL VAGLGADVEF
1351  GNGWNGLARY SYAGSKQYGN HSGRVGVGYR FLEHHHHHH*
```

961cL-ORF46.1

```
   1  ATGAAACACT TTCCATCCAA AGTACTGACC ACAGCCATCC TTGCCACTTT
  51  CTGTAGCGGC GCACTGGCAG CCACAAACGA CGACGATGTT AAAAAAGCTG
 101  CCACTGTGGC CATTGCTGCT GCCTACAACA ATGGCCAAGA AATCAACGGT
 151  TTCAAAGCTG GAGAGACCAT CTACGACATT GATGAAGACG GCACAATTAC
 201  CAAAAAAGAC GCAACTGCAG CCGATGTTGA AGCCGACGAC TTTAAAGGTC
 251  TGGGTCTGAA AAAAGTCGTG ACTAACCTGA CCAAAACCGT CAATGAAAAC
 301  AAACAAAACG TCGATGCCAA AGTAAAAGCT GCAGAATCTG AAATAGAAAA
 351  GTTAACAACC AAGTTAGCAG ACACTGATGC CGCTTTAGCA GATACTGATG
 401  CCGCTCTGGA TGCAACCACC AACGCCTTGA ATAAATTGGG AGAAAATATA
 451  ACGACATTTG CTGAAGAGAC TAAGACAAAT ATCGTAAAAA TTGATGAAAA
 501  ATTAGAAGCC GTGGCTGATA CCGTCGACAA GCATGCCGAA GCATTCAACG
 551  ATATCGCCGA TTCATTGGAT GAAACCAACA CTAAGGCAGA CGAAGCCGTC
 601  AAAACCGCCA ATGAAGCCAA ACAGACGGCC GAAGAAACCA AACAAAACGT
 651  CGATGCCAAA GTAAAAGCTG CAGAAACTGC AGCAGGCAAA GCCGAAGCTG
 701  CCGCTGGCAC AGCTAATACT GCAGCCGACA AGGCCGAAGC TGTCGCTGCA
 751  AAAGTTACCG ACATCAAAGC TGATATCGCT ACGAACAAAG ATAATATTGC
 801  TAAAAAAGCA AACAGTGCCG ACGTGTACAC CAGAGAAGAG TCTGACAGCA
 851  AATTTGTCAG AATTGATGGT CTGAACGCTA CTACCGAAAA ATTGGACACA
 901  CGCTTGGCTT CTGCTGAAAA ATCCATTGCC GATCACGATA CTCGCCTGAA
 951  CGGTTTGGAT AAAACAGTGT CAGACCTGCG CAAAGAAACC CGCCAAGGCC
1001  TTGCAGAACA AGCCGCGCTC TCCGGTCTGT TCCAACCTTA CAACGTGGGT
1051  GGATCCGGAG GAGGAGGATC AGATTTGGCA AACGATTCTT TTATCCGGCA
1101  GGTTCTCGAC CGTCAGCATT TCGAACCCGA CGGGAAATAC CACCTATTCG
1151  GCAGCAGGGG GGAACTTGCC GAGCGCAGCG GCCATATCGG ATTGGGAAAA
1201  ATACAAAGCC ATCAGTTGGG CAACCTGATG ATTCAACAGG CGGCCATTAA
1251  AGGAAATATC GGCTACATTG TCCGCTTTTC CGATCACGGG CACGAAGTCC
1301  ATTCCCCCTT CGACAACCAT GCCTCACATT CCGATTCTGA TGAAGCCGGT
1351  AGTCCCGTTG ACGGATTTAG CCTTTACCGC ATCCATTGGG ACGGATACGA
1401  ACACCATCCC GCCGACGGCT ATGACGGGCC ACAGGGCGGC GGCTATCCCG
```

```
1451  CTCCCAAAGG CGCGAGGGAT ATATACAGCT ACGACATAAA AGGCGTTGCC
1501  CAAAATATCC GCCTCAACCT GACCGACAAC CGCAGCACCG GACAACGGCT
1551  TGCCGACCGT TTCCACAATG CCGGTAGTAT GCTGACGCAA GGAGTAGGCG
1601  ACGGATTCAA ACGCGCCACC CGATACAGCC CCGAGCTGGA CAGATCGGGC
1651  AATGCCGCCG AAGCCTTCAA CGGCACTGCA GATATCGTTA AAAACATCAT
1701  CGGCGCGGCA GGAGAAATTG TCGGCGCAGG CGATGCCGTG CAGGGCATAA
1751  GCGAAGGCTC AAACATTGCT GTCATGCACG GCTTGGGTCT GCTTTCCACC
1801  GAAAACAAGA TGGCGCGCAT CAACGATTTG GCAGATATGG CGCAACTCAA
1851  AGACTATGCC GCAGCAGCCA TCCGCGATTG GGCAGTCCAA AACCCCAATG
1901  CCGCACAAGG CATAGAAGCC GTCAGCAATA TCTTTATGGC AGCCATCCCC
1951  ATCAAAGGGA TTGGAGCTGT TCGGGGAAAA TACGGCTTGG GCGGCATCAC
2001  GGCACATCCT ATCAAGCGGT CGCAGATGGG CGCGATCGCA TTGCCGAAAG
2051  GGAAATCCGC CGTCAGCGAC AATTTTGCCG ATGCGGCATA CGCCAAATAC
2101  CCGTCCCCTT ACCATTCCCG AAATATCCGT TCAAACTTGG AGCAGCGTTA
2151  CGGCAAAGAA AACATCACCT CCTCAACCGT GCCGCCGTCA AACGGCAAAA
2201  ATGTCAAACT GGCAGACCAA CGCCACCCGA AGACAGGCGT ACCGTTTGAC
2251  GGTAAAGGGT TTCCGAATTT TGAGAAGCAC GTGAAATATG ATACGTAACT
2301  CGAG

   1  MKHFPSKVLT TAILATFCSG ALAATNDDDV KKAATVAIAA AYNNGQEING
  51  FKAGETIYDI DEDGTITKKD ATAADVEADD FKGLGLKKVV TNLTKTVNEN
 101  KQNVDAKVKA AESEIEKLTT KLADTDAALA DTDAALDATT NALNKLGENI
 151  TTFAEETKTN IVKIDEKLEA VADTVDKHAE AFNDIADSLD ETNTKADEAV
 201  KTANEAKQTA EETKQNVDAK VKAAETAAGK AEAAAGTANT AADKAEAVAA
 251  KVTDIKADIA TNKDNIAKKA NSADVYTREE SDSKFVRIDG LNATTEKLDT
 301  RLASAEKSIA DHDTRLNGLD KTVSDLRKET RQGLAEQAAL SGLFQPYNVG
 351  GSGGGGSDLA NDSFIRQVLD RQHFEPDGKY HLFGSRGELA ERSGHIGLGK
 401  IQSHQLGNLM IQQAAIKGNI GYIVRFSDHG HEVHSPFDNH ASHSDSDEAG
 451  SPVDGFSLYR IHWDGYEHHP ADGYDGPQGG GYPAPKGARD IYSYDIKGVA
 501  QNIRLNLTDN RSTGQRLADR FHNAGSMLTQ GVGDGFKRAT RYSPELDRSG
 551  NAAEAFNGTA DIVKNIIGAA GEIVGAGDAV QGISEGSNIA VMHGLGLLST
 601  ENKMARINDL ADMAQLKDYA AAAIRDWAVQ NPNAAQGIEA VSNIFMAAIP
 651  IKGIGAVRGK YGLGGITAHP IKRSQMGAIA LPKGKSAVSD NFADAAYAKY
 701  PSPYHSRNIR SNLEQRYGKE NITSSTVPPS NGKNVKLADQ RHPKTGVPFD
 751  GKGFPNFEKH VKYDT*
```

961cL-741

```
   1  ATGAAACACT TTCCATCCAA AGTACTGACC ACAGCCATCC TTGCCACTTT
  51  CTGTAGCGGC GCACTGGCAG CCACAAACGA CGACGATGTT AAAAAAGCTG
 101  CCACTGTGGC CATTGCTGCT GCCTACAACA ATGGCCAAGA AATCAACGGT
 151  TTCAAAGCTG GAGAGACCAT CTACGACATT GATGAAGACG GCACAATTAC
 201  CAAAAAAGAC GCAACTGCAG CCGATGTTGA AGCCGACGAC TTTAAAGGTC
 251  TGGGTCTGAA AAAAGTCGTG ACTAACCTGA CCAAAACCGT CAATGAAAAC
 301  AAACAAAACG TCGATGCCAA AGTAAAAGCT GCAGAATCTG AAATAGAAAA
 351  GTTAACAACC AAGTTAGCAG ACACTGATGC CGCTTTAGCA GATACTGATG
 401  CCGCTCTGGA TGCAACCACC AACGCCTTGA ATAAATTGGG AGAAAATATA
 451  ACGACATTTG CTGAAGAGAC TAAGACAAAT ATCGTAAAAA TTGATGAAAA
 501  ATTAGAAGCC GTGGCTGATA CCGTCGACAA GCATGCCGAA GCATTCAACG
 551  ATATCGCCGA TTCATTGGAT GAAACCAACA CTAAGGCAGA CGAAGCCGTC
 601  AAAACCGCCA ATGAAGCCAA ACAGACGGCC GAAGAAACCA AACAAAACGT
 651  CGATGCCAAA GTAAAAGCTG CAGAAACTGC AGCAGGCAAA GCCGAAGCTG
 701  CCGCTGGCAC AGCTAATACT GCAGCCGACA AGGCCGAAGC TGTCGCTGCA
 751  AAAGTTACCG ACATCAAAGC TGATATCGCT ACGAACAAAG ATAATATTGC
 801  TAAAAAAGCA AACAGTGCCG ACGTGTACAC CAGAGAAGAG TCTGACAGCA
 851  AATTTGTCAG AATTGATGGT CTGAACGCTA CTACCGAAAA ATTGGACACA
 901  CGCTTGGCTT CTGCTGAAAA ATCCATTGCC GATCACGATA CTCGCCTGAA
 951  CGGTTTGGAT AAAACAGTGT CAGACCTGCG CAAAGAAACC CGCCAAGGCC
1001  TTGCAGAACA AGCCGCGCTC TCCGGTCTGT TCCAACCTTA CAACGTGGGT
1051  GGATCCGGAG GGGGTGGTGT CGCCGCCGAC ATCGGTGCGG GGCTTGCCGA
1101  TGCACTAACC GCACCGCTCG ACCATAAAGA CAAAGGTTTG CAGTCTTTGA
1151  CGCTGGATCA GTCCGTCAGG AAAAACGAGA AACTGAAGCT GGCGGCACAA
1201  GGTGCGGAAA AAACTTATGG AAACGGTGAC AGCCTCAATA CGGGCAAATT
1251  GAAGAACGAC AAGGTCAGCC GTTTCGACTT TATCCGCCAA ATCGAAGTGG
1301  ACGGGCAGCT CATTACCTTG GAGAGTGGAG AGTTCCAAGT ATACAAACAA
1351  AGCCATTCCG CCTTAACCGC CTTTCAGACC GAGCAAATAC AAGATTCGGA
1401  GCATTCCGGG AAGATGGTTG CGAAACGCCA GTTCAGAATC GGCGACATAG
```

```
1451  CGGGCGAACA TACATCTTTT GACAAGCTTC CCGAAGGCGG CAGGGCGACA
1501  TATCGCGGGA CGGCGTTCGG TTCAGACGAT GCCGGCGGAA AACTGACCTA
1551  CACCATAGAT TTCGCCGCCA AGCAGGGAAA CGGCAAAATC GAACATTTGA
1601  AATCGCCAGA ACTCAATGTC GACCTGGCCG CCGCCGATAT CAAGCCGGAT
1651  GGAAAACGCC ATGCCGTCAT CAGCGGTTCC GTCCTTTACA ACCAAGCCGA
1701  GAAAGGCAGT TACTCCCTCG GTATCTTTGG CGGAAAAGCC CAGGAAGTTG
1751  CCGGCAGCGC GGAAGTGAAA ACCGTAAACG GCATACGCCA TATCGGCCTT
1801  GCCGCCAAGC AACTCGAGCA CCACCACCAC CACCACTGA

   1  MKHFPSKVLT TAILATFCSG ALAATNDDDV KKAATVAIAA AYNNGQEING
  51  FKAGETIYDI DEDGTITKKD ATAADVEADD FKGLGLKKVV TNLTKTVNEN
 101  KQNVDAKVKA AESEIEKLTT KLADTDAALA DTDAALDATT NALNKLGENI
 151  TTFAEETKTN IVKIDEKLEA VADTVDKHAE AFNDIADSLD ETNTKADEAV
 201  KTANEAKQTA EETKQNVDAK VKAAETAAGK AEAAAGTANT AADKAEAVAA
 251  KVTDIKADIA TNKDNIAKKA NSADVYTREE SDSKFVRIDG LNATTEKLDT
 301  RLASAEKSIA DHDTRLNGLD KTVSDLRKET RQGLAEQAAL SGLFQPYNVG
 351  GSGGGGVAAD IGAGLADALT APLDHKDKGL QSLTLDQSVR KNEKLKLAAQ
 401  GAEKTYGNGD SLNTGKLKND KVSRFDFIRQ IEVDGQLITL ESGEFQVYKQ
 451  SHSALTAFQT EQIQDSEHSG KMVAKRQFRI GDIAGEHTSF DKLPEGGRAT
 501  YRGTAFGSDD AGGKLTYTID FAAKQGNGKI EHLKSPELNV DLAAADIKPD
 551  GKRHAVISGS VLYNQAEKGS YSLGIFGGKA QEVAGSAEVK TVNGIRHIGL
 601  AAKQLEHHHH HH*
```

961cL-983

```
   1  ATGAAACACT TTCCATCCAA AGTACTGACC ACAGCCATCC TTGCCACTTT
  51  CTGTAGCGGC GCACTGGCAG CCACAAACGA CGACGATGTT AAAAAAGCTG
 101  CCACTGTGGC CATTGCTGCT GCCTACAACA ATGGCCAAGA AATCAACGGT
 151  TTCAAAGCTG GAGAGACCAT CTACGACATT GATGAAGACG GCACAATTAC
 201  CAAAAAAGAC GCAACTGCAG CCGATGTTGA AGCCGACGAC TTTAAAGGTC
 251  TGGGTCTGAA AAAAGTCGTG ACTAACCTGA CCAAAACCGT CAATGAAAAC
 301  AAACAAAACG TCGATGCCAA AGTAAAAGCT GCAGAATCTG AAATAGAAAA
 351  GTTAACAACC AAGTTAGCAG ACACTGATGC CGCTTTAGCA GATACTGATG
 401  CCGCTCTGGA TGCAACCACC AACGCCTTGA ATAAATTGGG AGAAAATATA
 451  ACGACATTTG CTGAAGAGAC TAAGACAAAT ATCGTAAAAA TTGATGAAAA
 501  ATTAGAAGCC GTGGCTGATA CCGTCGACAA GCATGCCGAA GCATTCAACG
 551  ATATCGCCGA TTCATTGGAT GAAACCAACA CTAAGGCAGA CGAAGCCGTC
 601  AAAACCGCCA ATGAAGCCAA ACAGACGGCC GAAGAAACCA AACAAAACGT
 651  CGATGCCAAA GTAAAAGCTG CAGAAACTGC AGCAGGCAAA GCCGAAGCTG
 701  CCGCTGGCAC AGCTAATACT GCAGCCGACA AGGCCGAAGC TGTCGCTGCA
 751  AAAGTTACCG ACATCAAAGC TGATATCGCT ACGAACAAAG ATAATATTGC
 801  TAAAAAAGCA AACAGTGCCG ACGTGTACAC CAGAGAAGAG TCTGACAGCA
 851  AATTTGTCAG AATTGATGGT CTGAACGCTA CTACCGAAAA ATTGGACACA
 901  CGCTTGGCTT CTGCTGAAAA ATCCATTGCC GATCACGATA CTCGCCTGAA
 951  CGGTTTGGAT AAAACAGTGT CAGACCTGCG CAAAGAAACC CGCCAAGGCC
1001  TTGCAGAACA AGCCGCGCTC TCCGGTCTGT TCCAACCTTA CAACGTGGGT
1051  GGATCCGGCG GAGGCGGCAC TTCTGCGCCC GACTTCAATG CAGGCGGTAC
1101  CGGTATCGGC AGCAACAGCA GAGCAACAAC AGCGAAATCA GCAGCAGTAT
1151  CTTACGCCGG TATCAAGAAC GAAATGTGCA AAGACAGAAG CATGCTCTGT
1201  GCCGGTCGGG ATGACGTTGC GGTTACAGAC AGGGATGCCA AAATCAATGC
1251  CCCCCCCCCG AATCTGCATA CCGGAGACTT TCCAAACCCA AATGACGCAT
1301  ACAAGAATTT GATCAACCTC AAACCTGCAA TTGAAGCAGG CTATACAGGA
1351  CGCGGGGTAG AGGTAGGTAT CGTCGACACA GGCGAATCCG TCGGCAGCAT
1401  ATCCTTTCCC GAACTGTATG GCAGAAAAGA ACACGGCTAT AACGAAAATT
1451  ACAAAAACTA TACGGCGTAT ATGCGGAAGG AAGCGCCTGA AGACGGAGGC
1501  GGTAAAGACA TTGAAGCTTC TTTCGACGAT GAGGCCGTTA TAGAGACTGA
1551  AGCAAAGCCG ACGGATATCC GCCACGTAAA AGAAATCGGA CACATCGATT
1601  TGGTCTCCCA TATTATTGGC GGGCGTTCCG TGGACGGCAG ACCTGCAGGC
1651  GGTATTGCGC CCGATGCGAC GCTACACATA ATGAATACGA ATGATGAAAC
1701  CAAGAACGAA ATGATGGTTG CAGCCATCCG CAATGCATGG GTCAAGCTGG
1751  GCGAACGTGG CGTGCGCATC GTCAATAACA GTTTTGGAAC AACATCGAGG
1801  GCAGGCACTG CCGACCTTTT CCAAATAGCC AATTCGGAGG AGCAGTACCG
1851  CCAAGCGTTG CTCGACTATT CCGGCGGTGA TAAAACAGAC GAGGGTATCC
1901  GCCTGATGCA ACAGAGCGAT TACGGCAACC TGTCCTACCA CATCCGTAAT
1951  AAAAACATGC TTTTCATCTT TTCGACAGGC AATGACGCAC AAGCTCAGCC
2001  CAACACATAT GCCCTATTGC CATTTTATGA AAAAGACGCT CAAAAAGGCA
2051  TTATCACAGT CGCAGGCGTA GACCGCAGTG GAGAAAAGTT CAAACGGGAA
```

```
2101  ATGTATGGAG AACCGGGTAC AGAACCGCTT GAGTATGGCT CCAACCATTG
2151  CGGAATTACT GCCATGTGGT GCCTGTCGGC ACCCTATGAA GCAAGCGTCC
2201  GTTTCACCCG TACAAACCCG ATTCAAATTG CCGGAACATC CTTTTCCGCA
2251  CCCATCGTAA CCGGCACGGC GGCTCTGCTG CTGCAGAAAT ACCCGTGGAT
2301  GAGCAACGAC AACCTGCGTA CCACGTTGCT GACGACGGCT CAGGACATCG
2351  GTGCAGTCGG CGTGGACAGC AAGTTCGGCT GGGGACTGCT GGATGCGGGT
2401  AAGGCCATGA ACGGACCCGC GTCCTTTCCG TTCGGCGACT TTACCGCCGA
2451  TACGAAAGGT ACATCCGATA TTGCCTACTC CTTCCGTAAC GACATTTCAG
2501  GCACGGGCGG CCTGATCAAA AAAGGCGGCA GCCAACTGCA ACTGCACGGC
2551  AACAACACCT ATACGGGCAA AACCATTATC GAAGGCGGTT CGCTGGTGTT
2601  GTACGGCAAC AACAAATCGG ATATGCGCGT CGAAACCAAA GGTGCGCTGA
2651  TTTATAACGG GGCGGCATCC GGCGGCAGCC TGAACAGCGA CGGCATTGTC
2701  TATCTGGCAG ATACCGACCA ATCCGGCGCA AACGAAACCG TACACATCAA
2751  AGGCAGTCTG CAGCTGGACG GCAAAGGTAC GCTGTACACA CGTTTGGGCA
2801  AACTGCTGAA AGTGGACGGT ACGGCGATTA TCGGCGGCAA GCTGTACATG
2851  TCGGCACGCG GCAAGGGGGC AGGCTATCTC AACAGTACCG GACGACGTGT
2901  TCCCTTCCTG AGTGCCGCCA AAATCGGGCA GGATTATTCT TTCTTCACAA
2951  ACATCGAAAC CGACGGCGGC CTGCTGGCTT CCCTCGACAG CGTCGAAAAA
3001  ACAGCGGGCA GTGAAGGCGA CACGCTGTCC TATTATGTCC GTCGCGGCAA
3051  TGCGGCACGG ACTGCTTCGG CAGCGGCACA TTCCGCGCCC GCCGGTCTGA
3101  AACACGCCGT AGAACAGGGC GGCAGCAATC TGGAAAACCT GATGGTCGAA
3151  CTGGATGCCT CCGAATCATC CGCAACACCC GAGACGGTTG AAACTGCGGC
3201  AGCCGACCGC ACAGATATGC CGGGCATCCG CCCCTACGGC GCAACTTTCC
3251  GCGCAGCGGC AGCCGTACAG CATGCGAATG CCGCCGACGG TGTACGCATC
3301  TTCAACAGTC TCGCCGCTAC CGTCTATGCC GACAGTACCG CCGCCCATGC
3351  CGATATGCAG GGACGCCGCC TGAAAGCCGT ATCGGACGGG TTGGACCACA
3401  ACGGCACGGG TCTGCGCGTC ATCGCGCAAA CCCAACAGGA CGGTGGAACG
3451  TGGGAACAGG GCGGTGTTGA AGGCAAAATG CGCGGCAGTA CCCAAACCGT
3501  CGGCATTGCC GCGAAAACCG GCGAAAATAC GACAGCAGCC GCCACACTGG
3551  GCATGGGACG CAGCACATGG AGCGAAAACA GTGCAAATGC AAAAACCGAC
3601  AGCATTAGTC TGTTTGCAGG CATACGGCAC GATGCGGGCG ATATCGGCTA
3651  TCTCAAAGGC CTGTTCTCCT ACGGACGCTA CAAAAACAGC ATCAGCCGCA
3701  GCACCGGTGC GGACGAACAT GCGGAAGGCA GCGTCAACGG CACGCTGATG
3751  CAGCTGGGCG CACTGGGCGG TGTCAACGTT CCGTTTGCCG CAACGGGAGA
3801  TTTGACGGTC GAAGGCGGTC TGCGCTACGA CCTGCTCAAA CAGGATGCAT
3851  TCGCCGAAAA AGGCAGTGCT TTGGGCTGGA GCGGCAACAG CCTCACTGAA
3901  GGCACGCTGG TCGGACTCGC GGGTCTGAAG CTGTCGCAAC CCTTGAGCGA
3951  TAAAGCCGTC CTGTTTGCAA CGGCGGGCGT GGAACGCGAC CTGAACGGAC
4001  GCGACTACAC GGTAACGGGC GGCTTTACCG GCGCGACTGC AGCAACCGGC
4051  AAGACGGGGG CACGCAATAT GCCGCACACC CGTCTGGTTG CCGGCCTGGG
4101  CGCGGATGTC GAATTCGGCA ACGGCTGGAA CGGCTTGGCA CGTTACAGCT
4151  ACGCCGGTTC CAAACAGTAC GGCAACCACA GCGGACGAGT CGGCGTAGGC
4201  TACCGGTTCT GACTCGAG

   1  MKHFPSKVLT TAILATFCSG ALAATNDDDV KKAATVAIAA AYNNGQEING
  51  FKAGETIYDI DEDGTITKKD ATAADVEADD FKGLGLKKVV TNLTKTVNEN
 101  KQNVDAKVKA AESEIEKLTT KLADTDAALA DTDAALDATT NALNKLGENI
 151  TTFAEETKTN IVKIDEKLEA VADTVDKHAE AFNDIADSLD ETNTKADEAV
 201  KTANEAKQTA EETKQNVDAK VKAAETAAGK AEAAAGTANT AADKAEAVAA
 251  KVTDIKADIA TNKDNIAKKA NSADVYTREE SDSKFVRIDG LNATTEKLDT
 301  RLASAEKSIA DHDTRLNGLD KTVSDLRKET RQGLAEQAAL SGLFQPYNVG
 351  GSGGGGTSAP DFNAGGTGIG SNSRATTAKS AAVSYAGIKN EMCKDRSMLC
 401  AGRDDVAVTD RDAKINAPPP NLHTGDFPNP NDAYKNLINL KPAIEAGYTG
 451  RGVEVGIVDT GESVGSISFP ELYGRKEHGY NENYKNYTAY MRKEAPEDGG
 501  GKDIEASFDD EAVIETEAKP TDIRHVKEIG HIDLVSHIIG GRSVDGRPAG
 551  GIAPDATLHI MNTNDETKNE MMVAAIRNAW VKLGERGVRI VNNSFGTTSR
 601  AGTADLFQIA NSEEQYRQAL LDYSGGDKTD EGIRLMQQSD YGNLSYHIRN
 651  KNMLFIFSTG NDAQAQPNTY ALLPFYEKDA QKGIITVAGV DRSGEKFKRE
 701  MYGEPGTEPL EYGSNHCGIT AMWCLSAPYE ASVRFTRTNP IQIAGTSFSA
 751  PIVTGTAALL LQKYPWMSND NLRTTLLTTA QDIGAVGVDS KFGWGLLDAG
 801  KAMNGPASFP FGDFTADTKG TSDIAYSFRN DISGTGGLIK KGGSQLQLHG
 851  NNTYTGKTII EGGSLVLYGN NKSDMRVETK GALIYNGAAS GGSLNSDGIV
 901  YLADTDQSGA NETVHIKGSL QLDGKGTLYT RLGKLLKVDG TAIIGGKLYM
 951  SARGKGAGYL NSTGRRVPFL SAAKIGQDYS FFTNIETDGG LLASLDSVEK
1001  TAGSEGDTLS YYVRRGNAAR TASAAAHSAP AGLKHAVEQG GSNLENLMVE
1051  LDASESSATP ETVETAAADR TDMPGIRPYG ATFRAAAAVQ HANAADGVRI
1101  FNSLAATVYA DSTAAHADMQ GRRLKAVSDG LDHNGTGLRV IAQTQQDGGT
```

```
1151   WEQGGVEGKM RGSTQTVGIA AKTGENTTAA ATLGMGRSTW SENSANAKTD
1201   SISLFAGIRH DAGDIGYLKG LFSYGRYKNS ISRSTGADEH AEGSVNGTLM
1251   QLGALGGVNV PFAATGDLTV EGGLRYDLLK QDAFAEKGSA LGWSGNSLTE
1301   GTLVGLAGLK LSQPLSDKAV LFATAGVERD LNGRDYTVTG GFTGATAATG
1351   KTGARNMPHT RLVAGLGADV EFGNGWNGLA RYSYAGSKQY GNHSGRVGVG
1401   YRF*
```

It will be understood that the invention has been described by way of example only and modifications may be made whilst remaining within the scope and spirit of the invention. For instance, the use of proteins from other strains is envisaged [*e.g.* see WO00/66741 for polymorphic sequences for ORF4, ORF40, ORF46, 225, 235, 287, 519, 726, 919 and 953].

EXPERIMENTAL DETAILS

FPLC protein purification

The following table summarises the FPLC protein purification that was used:

Protein	PI	Column	Buffer	pH	Protocol
121.1^{untagged}	6.23	Mono Q	Tris	8.0	A
128.1^{untagged}	5.04	Mono Q	Bis-Tris propane	6.5	A
406.1L	7.75	Mono Q	Diethanolamine	9.0	B
576.1L	5.63	Mono Q	Tris	7.5	B
593^{untagged}	8.79	Mono S	Hepes	7.4	A
726^{untagged}	4.95	Hi-trap S	Bis-Tris	6.0	A
919^{untagged}	10.5(-leader)	Mono S	Bicine	8.5	C
919Lorf4	10.4(-leader)	Mono S	Tris	8.0	B
920L	6.92(-leader)	Mono Q	Diethanolamine	8.5	A
953L	7.56(-leader)	Mono S	MES	6.6	D
982^{untagged}	4.73	Mono Q	Bis-Tris propane	6.5	A
919-287	6.58	Hi-trap Q	Tris	8.0	A
953-287	4.92	Mono Q	Bis-Tris propane	6.2	A

Buffer solutions included 20-120 mM NaCl, 5.0 mg/ml CHAPS and 10% v/v glycerol. The dialysate was centrifuged at 13000*g* for 20 min and applied to either a mono Q or mono S FPLC ion-exchange resin. Buffer and ion exchange resins were chosen according to the pI of the protein of interest and the recommendations of the FPLC protocol manual [Pharmacia: *FPLC Ion Exchange and Chromatofocussing; Principles and Methods.* Pharmacia

Publication]. Proteins were eluted using a step-wise NaCl gradient. Purification was analysed by SDS-PAGE and protein concentration determined by the Bradford method.

The letter in the 'protocol' column refers to the following:

FPLC-A: Clones 121.1, 128.1, 593, 726, 982, periplasmic protein 920L and hybrid proteins 919-287, 953-287 were purified from the soluble fraction of *E.coli* obtained after disruption of the cells. Single colonies harbouring the plasmid of interest were grown overnight at 37°C in 20 ml of LB/Amp (100 µg/ml) liquid culture. Bacteria were diluted 1:30 in 1.0 L of fresh medium and grown at either 30°C or 37°C until the OD_{550} reached 0.6-08. Expression of recombinant protein was induced with IPTG at a final concentration of 1.0 mM. After incubation for 3 hours, bacteria were harvested by centrifugation at 8000*g* for 15 minutes at 4°C. When necessary cells were stored at -20°C. All subsequent procedures were performed on ice or at 4°C. For cytosolic proteins (121.1, 128.1, 593, 726 and 982) and periplasmic protein 920L, bacteria were resuspended in 25 ml of PBS containing complete protease inhibitor (Boehringer-Mannheim). Cells were lysed by by sonication using a Branson Sonifier 450. Disrupted cells were centrifuged at 8000*g* for 30 min to sediment unbroken cells and inclusion bodies and the supernatant taken to 35% v/v saturation by the addition of 3.9 M $(NH_4)_2SO_4$. The precipitate was sedimented at 8000*g* for 30 minutes. The supernatant was taken to 70% v/v saturation by the addition of 3.9 M $(NH_4)_2SO_4$ and the precipitate collected as above. Pellets containing the protein of interest were identified by SDS-PAGE and dialysed against the appropriate ion-exchange buffer (see below) for 6 hours or overnight. The periplasmic fraction from *E.coli* expressing 953L was prepared according to the protocol of Evans *et. al.* [*Infect.Immun.* (1974) 10:1010-1017] and dialysed against the appropriate ion-exchange buffer. Buffer and ion exchange resin were chosen according to the pI of the protein of interest and the recommendations of the FPLC protocol manual (Pharmacia). Buffer solutions included 20 mM NaCl, and 10% (v/v) glycerol. The dialysate was centrifuged at 13000*g* for 20 min and applied to either a mono Q or mono S FPLC ion-exchange resin. Buffer and ion exchange resin were chosen according to the pI of the protein of interest and the recommendations of the FPLC protocol manual (Pharmacia). Proteins were eluted from the ion-exchange resin using either step-wise or continuous NaCl gradients. Purification was analysed by SDS-PAGE and protein concentration determined by Bradford method. Cleavage of the leader peptide of periplasmic proteins was demonstrated by sequencing the NH_2-terminus (see below).

FPLC-B: These proteins were purified from the membrane fraction of *E.coli*. Single colonies harbouring the plasmid of interest were grown overnight at 37°C in 20 ml of LB/Amp (100 µg/ml) liquid culture. Bacteria were diluted 1:30 in 1.0 L of fresh medium. Clones 406.1L and 919LOrf4 were grown at 30°C and Orf25L and 576.1L at 37°C until the OD_{550} reached 0.6-0.8. In the case of 919LOrf4, growth at 30°C was essential since expression of recombinant protein at 37°C resulted in lysis of the cells. Expression of recombinant protein was induced with IPTG at a final concentration of 1.0 mM. After incubation for 3 hours, bacteria were harvested by centrifugation at 8000*g* for 15 minutes at 4°C. When necessary cells were stored at -20 °C. All subsequent procedures were performed at 4°C. Bacteria were resuspended in 25 ml of PBS containing complete protease inhibitor (Boehringer-Mannheim) and lysed by osmotic shock with 2-3 passages through a French Press. Unbroken cells were removed by centrifugation at 5000*g* for 15 min and membranes precipitated by centrifugation at 100000*g* (Beckman Ti50, 38000rpm) for 45 minutes. A Dounce homogenizer was used to re-suspend the membrane pellet in 7.5 ml of 20 mM Tris-HCl (pH 8.0), 1.0 M NaCl and complete protease inhibitor. The suspension was mixed for 2-4 hours, centrifuged at 100000*g* for 45 min and the pellet resuspended in 7.5 ml of 20mM Tris-HCl (pH 8.0), 1.0M NaCl, 5.0mg/ml CHAPS, 10% (v/v) glycerol and complete protease inhibitor. The solution was mixed overnight, centrifuged at 100000*g* for 45 minutes and the supernatant dialysed for 6 hours against an appropriately selected buffer. In the case of Orf25.L, the pellet obtained after CHAPS extraction was found to contain the recombinant protein. This fraction, without further purification, was used to immunise mice.

FPLC-C: Identical to FPLC-A, but purification was from the soluble fraction obtained after permeabilising *E.coli* with polymyxin B, rather than after cell disruption.

FPLC-D: A single colony harbouring the plasmid of interest was grown overnight at 37°C in 20 ml of LB/Amp (100 µg/ml) liquid culture. Bacteria were diluted 1:30 in 1.0 L of fresh medium and grown at 30°C until the OD_{550} reached 0.6-0.8. Expression of recombinant protein was induced with IPTG at a final concentration of 1.0mM. After incubation for 3 hours, bacteria were harvested by centrifugation at 8000*g* for 15 minutes at 4°C. When necessary cells were stored at -20 °C. All subsequent procedures were performed on ice or at 4°C. Cells were resuspended in 20mM Bicine (pH 8.5), 20mM NaCl, 10% (v/v) glycerol, complete protease inhibitor (Boehringer-Mannheim) and disrupted using a Branson Sonifier 450. The sonicate was centrifuged at 8000*g* for 30 min to sediment unbroken cells and

inclusion bodies. The recombinant protein was precipitated from solution between 35% v/v and 70% v/v saturation by the addition of 3.9M $(NH_4)_2SO_4$. The precipitate was sedimented at 8000*g* for 30 minutes, resuspended in 20 mM Bicine (pH 8.5), 20 mM NaCl, 10% (v/v) glycerol and dialysed against this buffer for 6 hours or overnight. The dialysate was centrifuged at 13000*g* for 20 min and applied to the FPLC resin. The protein was eluted from the column using a step-wise NaCl gradients. Purification was analysed by SDS-PAGE and protein concentration determined by Bradford method.

Cloning strategy and oligonucleotide design

Genes coding for antigens of interest were amplified by PCR, using oligonucleotides designed on the basis of the genomic sequence of *N. meningitidis* B MC58. Genomic DNA from strain 2996 was always used as a template in PCR reactions, unless otherwise specified, and the amplified fragments were cloned in the expression vector pET21b+ (Novagen) to express the protein as C-terminal His-tagged product, or in pET-24b+(Novagen) to express the protein in 'untagged' form (*e.g.* ΔG 287K).

Where a protein was expressed without a fusion partner and with its own leader peptide (if present), amplification of the open reading frame (ATG to STOP codons) was performed.

Where a protein was expressed in 'untagged' form, the leader peptide was omitted by designing the 5'-end amplification primer downstream from the predicted leader sequence.

The melting temperature of the primers used in PCR depended on the number and type of hybridising nucleotides in the whole primer, and was determined using the formulae:

$T_{m1} = 4\ (G+C) + 2\ (A+T)$ (tail excluded)

$T_{m2} = 64.9 + 0.41\ (\%\ GC) - 600/N$ (whole primer)

The melting temperatures of the selected oligonucleotides were usually 65-70°C for the whole oligo and 50-60°C for the hybridising region alone.

Oligonucleotides were synthesised using a Perkin Elmer 394 DNA/RNA Synthesizer, eluted from the columns in 2.0ml NH_4OH, and deprotected by 5 hours incubation at 56°C. The oligos were precipitated by addition of 0.3M Na-Acetate and 2 volumes ethanol. The samples were centrifuged and the pellets resuspended in water.

		Sequences	Restriction site
Orf1L	Fwd	CGCGGATCCGCTAGC-AAAACAACCGACAAACGG	NheI
	Rev	CCCGCTCGAG-TTACCAGCGGTAGCCTA	XhoI
Orf1	Fwd	CTAGCTAGC-GGACACACTTATTTCGGCATC	NheI
	Rev	CCCGCTCGAG- TTACCAGCGGTAGCCTAATTTG	XhoI
Orf1LOmpA	Fwd		NdeI-(NheI)
	Rev	CCCGCTCGAG-	XhoI
Orf4L	Fwd	CGCGGATCCCATATG-AAAACCTTCTTCAAAACC	NdeI
	Rev	CCCGCTCGAG-TTATTTGGCTGCGCCTTC	XhoI
Orf7-1L	Fwd	GCGGCATTAAT-ATGTTGAGAAAATTGTTGAAATGG	AseI
	Rev	GCGGCCTCGAG-TTATTTTTTCAAAATATATTTGC	XhoI
Orf9-1L	Fwd	GCGGCCATATG-TTACCTAACCGTTTCAAAATGT	NdeI
	Rev	GCGGCCTCGAG-TTATTTCCGAGGTTTTCGGG	XhoI
Orf23L	Fwd	CGCGGATCCCATATG-ACACGCTTCAAATATTC	NdeI
	Rev	CCCGCTCGAG-TTATTTAAACCGATAGGTAAA	XhoI
Orf25-1 His	Fwd	CGCGGATCCCATATG-GGCAGGGAAGAACCGC	NdeI
	Rev	GCCCAAGCTT-ATCGATGGAATAGCCGCG	HindIII
Orf29-1 b-His (MC58)	Fwd	CGCGGATCCGCTAGC-AACGGTTTGGATGCCCG	NheI
	Rev	CCCGCTCGAG-TTTGTCTAAGTTCCTGATAT CCCGCTCGAG-ATTCCCACCTGCCATC	XhoI
Orf29-1 b-L (MC58)	Fwd	CGCGGATCCGCTAGC-ATGAATTTGCCTATTCAAAAAT	NheI
	Rev	CCCGCTCGAG-TTAATTCCCACCTGCCATC	XhoI
Orf29-1 c-His (MC58)	Fwd	CGCGGATCCGCTAGC-ATGAATTTGCCTATTCAAAAAT	NheI
	Rev	CCCGCTCGAG-TTGGACGATGCCCGCGA	XhoI
Orf29-1 c-L (MC58)	Fwd	CGCGGATCCGCTAGC-ATGAATTTGCCTATTCAAAAAT	NheI
	Rev	CCCGCTCGAG-TTATTGGACGATGCCCGC	XhoI
Orf25L	Fwd	CGCGGATCCCATATG-TATCGCAAACTGATTGC	NdeI
	Rev	CCCGCTCGAG-CTAATCGATGGAATAGCC	XhoI
Orf37L	Fwd	CGCGGATCCCATATG-AAACAGACAGTCAAATG	NdeI
	Rev	CCCGCTCGAG-TCAATAACCCGCCTTCAG	XhoI
Orf38L	Fwd	CGCGGATCCCATATG- TTACGTTTGACTGCTTTAGCCGTATGCACC	NdeI
	Rev	CCCGCTCGAG- TTATTTTGCCGCGTTAAAAGCGTCGGCAAC	XhoI
Orf40L	Fwd	CGCGGATCCCATATG-AACAAAATATACCGCAT	NdeI
	Rev	CCCGCTCGAG-TTACCACTGATAACCGAC	XhoI
Orf40.2-His	Fwd	CGCGGATCCCATATG-ACCGATGACGACGATTTAT	NdeI
	Rev	GCCCAAGCTT-CCACTGATAACCGACAGA	HindIII
Orf40.2L	Fwd	CGCGGATCCCATATG-AACAAAATATACCGCAT	NdeI
	Rev	GCCCAAGCTT-TTACCACTGATAACCGAC	HindIII
Orf46-2L	Fwd	GGGAATTCCATATG-GGCATTTCCCGCAAAATATC	NdeI
	Rev	CCCGCTCGAG-TTATTTACTCCTATAACGAGGTCTCTTAAC	XhoI
Orf46-2	Fwd	GGGAATTCCATATG-TCAGATTTGGCAAACGATTCTT	NdeI
	Rev	CCCGCTCGAG-TTATTTACTCCTATAACGAGGTCTCTTAAC	XhoI
Orf46.1L	Fwd	GGGAATTCCATATG-GGCATTTCCCGCAAAATATC	NdeI

	Rev	CCCGCTCGAG-TTACGTATCATATTTCACGTGC	XhoI
orf46. (His-GST)	Fwd	GGGAATTCCATATGCACGTGAAATATGATACGAAG	BamHI-NdeI
	Rev	CCCGCTCGAGTTTACTCCTATAACGAGGTCTCTTAAC	XhoI
orf46.1-His	Fwd	GGGAATTCCATATGTCAGATTTGGCAAACGATTCTT	NdeI
	Rev	CCCGCTCGAGCGTATCATATTTCACGTGC	XhoI
orf46.2-His	Fwd	GGGAATTCCATATGTCAGATTTGGCAAACGATTCTT	NdeI
	Rev	CCCGCTCGAGTTTACTCCTATAACGAGGTCTCTTAAC	XhoI
Orf65-1-(His/GST)	Fwd	CGCGGATCCCATATG-CAAAATGCGTTCAAAATCCC	BamHI-NdeI
(MC58)	Rev	CGCGGATCCCATATG-AACAAAATATACCGCAT CCCGCTCGAG –TTTGCTTTCGATAGAACGG	XhoI
Orf72-1L	Fwd	GCGGCCATATG-GTCATAAAATATACAAATTTGAA	NdeI
	Rev	GCGGCCTCGAG-TTAGCCTGAGACCTTTGCAAATT	XhoI
Orf76-1L	Fwd	GCGGCCATATG-AAACAGAAAAAAACCGCTG	NdeI
	Rev	GCGGCCTCGAG-TTACGGTTTGACACCGTTTTC	XhoI
Orf83.1L	Fwd	CGCGGATCCCATATG-AAAACCCTGCTCCTC	NdeI
	Rev	CCCGCTCGAG-TTATCCTCCTTTGCGGC	XhoI
Orf85-2L	Fwd	GCGGCCATATG-GCAAAAATGATGAAATGGG	NdeI
	Rev	GCGGCCTCGAG-TTATCGGCGCGGCGGGCC	XhoI
Orf91L (MC58)	Fwd	GCGGCCATATGAAAAAATCCTCCCTCATCA	NdeI
	Rev	GCGGCCTCGAGTTATTTGCCGCCGTTTTTGGC	XhoI
Orf91-His(MC58)	Fwd	GCGGCCATATGGCCCCTGCCGACGCGGTAAG	NdeI
	Rev	GCGGCCTCGAGTTTGCCGCCGTTTTTGGCTTTC	XhoI
Orf97-1L	Fwd	GCGGCCATATG-AAACACATACTCCCCCTGA	NdeI
	Rev	GCGGCCTCGAG-TTATTCGCCTACGGTTTTTTG	XhoI
Orf119L (MC58)	Fwd	GCGGCCATATGATTTACATCGTACTGTTTC	NdeI
	Rev	GCGGCCTCGAGTTAGGAGAACAGGCGCAATGC	XhoI
Orf119-His(MC58)	Fwd	GCGGCCATATGTACAACATGTATCAGGAAAAC	NdeI
	Rev	GCGGCCTCGAGGGAGAACAGGCGCAATGCGG	XhoI
Orf137.1 (His-GST) (MC58)	Fwd	CGCGGATCCGCTAGCTGCGGCACGGCGGG	BamHI-NheI
	Rec	CCCGCTCGAGATAACGGTATGCCGCCAG	XhoI
Orf143-1L	Fwd	CGCGGATCCCATATG-GAATCAACACTTTCAC	NdeI
	Rev	CCCGCTCGAG-TTACACGCGGTTGCTGT	XhoI
008	Fwd	CGCGGATCCCATATG-AACAACAGACATTTTG	NdeI
	Rev	CCCGCTCGAG-TTACCTGTCCGGTAAAAG	XhoI
050-1(48)	Fwd	CGCGGATCCGCTAGC-ACCGTCATCAAACAGGAA	NheI
	Rev	CCCGCTCGAG-TCAAGATTCGACGGGGA	XhoI
105	Fwd	CGCGGATCCCATATG-TCCGCAAACGAATACG	NdeI
	Rev	CCCGCTCGAG-TCAGTGTTCTGCCAGTTT	XhoI
111L	Fwd	CGCGGATCCCATATG-CCGTCTGAAACACG	NdeI
	Rev	CCCGCTCGAG-TTAGCGGAGCAGTTTTTC	XhoI
117-1	Fwd	CGCGGATCCCATATG-ACCGCCATCAGCC	NdeI
	Rev	CCCGCTCGAG-TTAAAGCCGGGTAACGC	XhoI
121-1	Fwd	GCGGCCATATG-GAAACACAGCTTTACATCGG	NdeI
	Rev	GCGGCCTCGAG-TCAATAATAATATCCCGCG	XhoI

122-1	Fwd	GCGGCCATATG-ATTAAAATCCGCAATATCC	NdeI
	Rev	GCGGCCTCGAG-TTAAATCTTGGTAGATTGGATTTGG	XhoI
128-1	Fwd	GCGGCCATATG-ACTGACAACGCACTGCTCC	NdeI
	Rev	GCGGCCTCGAG-TCAGACCGCGTTGTCGAAAC	XhoI
148	Fwd	CGCGGATCCCATATG-GCGTTAAAAACATCAAA	NdeI
	Rev	CCCGCTCGAG-TCAGCCCTTCATACAGC	XhoI
149.1L (MC58)	Fwd	GCGGCATTAATGGCACAAACTACACTCAAACC	AseI
	Rev	GCGGCCTCGAGTTAAAACTTCACGTTCACGCCG	XhoI
149.1-His(MC58)	Fwd	GCGGCATTAATGCATGAAACTGAGCAATCGGTGG	AseI
	Rev	GCGGCCTCGAGAAACTTCACGTTCACGCCGCCGGTAAA	XhoI
205 (His-GST) (MC58)	Fwd	CGCGGATCCCATATGGGCAAATCCGAAAATACG	BamHI-NdeI
	Rev	CCCGCTCGAGATAATGGCGGCGGCGG	XhoI
206L	Fwd	CGCGGATCCCATATG-TTTCCCCCCGACAA	NdeI
	Rev	CCCGCTCGAG-TCATTCTGTAAAAAAAGTATG	XhoI
214 (His-GST) (MC58)	Fwd	CGCGGATCCCATATGCTTCAAAGCGACAGCAG	BamHI-NdeI
	Rev	CCCGCTCGAGTTCGGATTTTTGCGTACTC	XhoI
216	Fwd	CGCGGATCCCATATG-GCAATGGCAGAAAACG	NdeI
	Rev	CCCGCTCGAG-CTATACAATCCGTGCCG	XhoI
225-1L	Fwd	CGCGGATCCCATATG-GATTCTTTTTTCAAACC	NdeI
	Rev	CCCGCTCGAG-TCAGTTCAGAAAGCGGG	XhoI
235L	Fwd	CGCGGATCCCATATG-AAACCTTTGATTTTAGG	NdeI
	Rev	CCCGCTCGAG-TTATTTGGGCTGCTCTTC	XhoI
243	Fwd	CGCGGATCCCATATG-GTAATCGTCTGGTTG	NdeI
	Rev	CCCGCTCGAG-CTACGACTTGGTTACCG	XhoI
247-1L	Fwd	GCGGCCATATG-AGACGTAAAATGCTAAAGCTAC	NdeI
	Rev	GCGGCCTCGAG-TCAAAGTGTTCTGTTTGCGC	XhoI
264-His	Fwd	GCCGCCATATG-TTGACTTTAACCCGAAAAA	NdeI
	Rev	GCCGCCTCGAG-GCCGGCGGTCAATACCGCCCGAA	XhoI
270 (His-GST) (MC58)	Fwd	CGCGGATCCCATATGGCGCAATGCGATTTGAC	BamHI-NdeI
	Rev	CCCGCTCGAGTTCGGCGGTAAATGCCG	XhoI
274L	Fwd	GCGGCCATATG-GCGGGGCCGATTTTTGT	NdeI
	Rev	GCGGCCTCGAG-TTATTTGCTTTCAGTATTATTG	XhoI
283L	Fwd	GCGGCCATATG-AACTTTGCTTTATCCGTCA	NdeI
	Rev	GCGGCCTCGAG-TTAACGGCAGTATTTGTTTAC	XhoI
285-His	Fwd	CGCGGATCCCATATGGGTTTGCGCTTCGGGC	BamHI
	Rev	GCCCAAGCTTTTTTCCTTTGCCGTTTCCG	HindIII
286-His (MC58)	Fwd	CGCGGATCCCATATG-GCCGACCTTTCCGAAAA	NdeI
	Rev	CCCGCTCGAG-GAAGCGCGTTCCCAAGC	XhoI
286L (MC58)	Fwd	CGCGGATCCCATATG-CACGACACCCGTAC	NdeI
	Rev	CCCGCTCGAG-TTAGAAGCGCGTTCCCAA	XhoI
287L	Fwd	CTAGCTAGC-TTTAAACGCAGCGTAATCGCAATGG	NheI
	Rev	CCCGCTCGAG-TCAATCCTGCTCTTTTTTGCC	XhoI

287	Fwd	CTAGCTAGC-GGGGGCGGCGGTGGCG	NheI
	Rev	CCCGCTCGAG-TCAATCCTGCTCTTTTTTGCC	XhoI
287LOrf4	Fwd	CTAGCTAGCGCTCATCCTCGCCGCC-TGCGGGGGCGGCGGT	NheI
	Rev	CCCGCTCGAG-TCAATCCTGCTCTTTTTTGCC	XhoI
287-fu	Fwd	CGGGGATCC-GGGGGCGGCGGTGGCG	BamHI
	Rev	CCCGCTCGAG-TCAATCCTGCTCTTTTTTGCC	XhoI
287-His	Fwd	CTAGCTAGC-GGGGGCGGCGGTGGCG	NheI
	Rev	CCCGCTCGAG-ATCCTGCTCTTTTTTGCC *	XhoI
287-His(2996)	Fwd	CTAGCTAGC-TGCGGGGGCGGCGGTGGCG	NheI
	Rev	CCCGCTCGAG-ATCCTGCTCTTTTTTGCC	XhoI
Δ1 287-His	Fwd	CGCGGATCCGCTAGC-CCCGATGTTAAATCGGC §	NheI
Δ2 287-His	Fwd	CGCGGATCCGCTAGC-CAAGATATGGCGGCAGT §	NheI
Δ3 287-His	Fwd	CGCGGATCCGCTAGC-GCCGAATCCGCAAATCA §	NheI
Δ4 287-His	Fwd	CGCGCTAGC-GGAAGGGTTGATTTGGCTAATGG §	NheI
Δ4 287MC58-His	Fwd	CGCGCTAGC-GGAAGGGTTGATTTGGCTAATGG §	NheI
287a-His	Fwd	CGCCATATG-TTTAAACGCAGCGTAATCGC	NdeI
	Rev	CCCGCTCGAG-AAAATTGCTACCGCCATTCGCAGG	XhoI
287b-His	Fwd	CGCCATATG-GGAAGGGTTGATTTGGCTAATGG	NdeI
287b-2996-His	Rev	CCCGCTCGAG-CTTGTCTTTATAAATGATGACATATTTG	XhoI
287b-MC58-His	Rev	CCCGCTCGAG-TTTATAAAAGATAATATATTGATTGATTCC	XhoI
287c-2996-His	Fwd	CGCGCTAGC-ATGCCGCTGATTCCCGTCAATC §	NheI
'287^{untagged}'(2996)	Fwd	CTAGCTAGC-GGGGGCGGCGGTGGCG	NheI
	Rev	CCCGCTCGAG-TCAATCCTGCTCTTTTTTGCC	XhoI
ΔG287-His *	Fwd	CGCGGATCCGCTAGC-CCCGATGTTAAATCGGC	NheI
	Rev	CCCGCTCGAG-ATCCTGCTCTTTTTTGCC	XhoI
ΔG287K(2996)	Fwd	CGCGGATCCGCTAGC-CCCGATGTTAAATCGGC	NheI
	Rev	CCCGCTCGAG-TCAATCCTGCTCTTTTTTGCC	XhoI
ΔG 287-L	Fwd	CGCGGATCCGCTAGC-TTTGAACGCAGTGTGATTGCAATGGCTTGTATTTTTGCC CTTTCAGCCTGT TCGCCCGATGTTAAATCGGCG	NheI
	Rev	CCCGCTCGAG-TCAATCCTGCTCTTTTTTGCC	XhoI
ΔG 287-Orf4L	Fwd	CGCGGATCCGCTAGC-AAAACCTTCTTCAAAACCCTTTCCGCCGCCGCACTCGCG CTCATCCTCGCCGCCTGC TCGCCCGATGTTAAATCG	NheI
	Rev	CCCGCTCGAG-TCAATCCTGCTCTTTTTTGCC	XhoI
292L	Fwd	CGCGGATCCCATATG-AAAACCAAGTTAATCAAA	NdeI
	Rev	CCCGCTCGAG-TTATTGATTTTTGCGGATGA	XhoI
308-1	Fwd	CGCGGATCCCATATG-TTAAATCGGGTATTTTATC	NdeI
	Rev	CCCGCTCGAG-TTAATCCGCCATTCCCTG	XhoI
401L	Fwd	GCGGCCATATG-AAATTACAACAATTGGCTG	NdeI
	Rev	GCGGCCTCGAG-TTACCTTACGTTTTTCAAAG	XhoI
406L	Fwd	CGCGGATCCCATATG-CAAGCACGGCTGCT	NdeI
	Rev	CCCGCTCGAG-TCAAGGTTGTCCTTGTCTA	XhoI
502-1L	Fwd	CGCGGATCCCATATG-ATGAAACCGCACAAC	NdeI
	Rev	CCCGCTCGAG-TCAGTTGCTCAACACGTC	XhoI

502-A (His-GST)	Fwd	CGCGGATCCCATATGGTAGACGCGCTTAAGCA	BamHI-NdeI
	Rev	CCCGCTCGAGAGCTGCATGGCGGCG	XhoI
503-1L	Fwd	CGCGGATCCCATATG-GCACGGTCGTTATAC	NdeI
	Rev	CCCGCTCGAG-CTACCGCGCATTCCTG	XhoI
519-1L	Fwd	GCGGCCATATG-GAATTTTTCATTATCTTGTT	NdeI
	Rev	GCGGCCTCGAG-TTATTTGGCGGTTTTGCTGC	XhoI
525-1L	Fwd	GCGGCCATATG-AAGTATGTCCGGTTATTTTTC	NdeI
	Rev	GCGGCCTCGAG-TTATCGGCTTGTGCAACGG	XhoI
529-(His/GST) (MC58)	Fwd	CGCGGATCCGCTAGC-TCCGGCAGCAAAACCGA	Bam HI-NheI
	Rev	GCCCAAGCTT-ACGCAGTTCGGAATGGAG	HindIII
552L	Fwd	GCCGCCATATGTTGAATATTAAACTGAAAACCTTG	NdeI
	Rev	GCCGCCTCGAGTTATTTCTGATGCCTTTTCCC	XhoI
556L	Fwd	GCCGCCATATGGACAATAAGACCAAACTG	NdeI
	Rev	GCCGCCTCGAGTTAACGGTGCGGACGTTTC	XhoI
557L	Fwd	CGCGGATCCCATATG-AACAAACTGTTTCTTAC	NdeI
	Rev	CCCGCTCGAG-TCATTCCGCCTTCAGAAA	XhoI
564ab-(His/GST) (MC58)	Fwd	CGCGGATCCCATATG-CAAGGTATCGTTGCCGACAAATCCGCACCT	BamHI-NdeI
	Rev	CCCGCTCGAG-AGCTAATTGTGCTTGGTTTGCAGATAGGAGTT	XhoI
564abL (MC58)	Fwd	CGCGGATCCCATATG-AACCGCACCCTGTACAAAGTTGTATTTAACAAACATC	NdeI
	Rev	CCCGCTCGAG-TTAAGCTAATTGTGCTTGGTTTGCAGATAGGAGTT	XhoI
564b-(His/GST)(MC58)	Fwd	CGCGGATCCCATATG-ACGGGAGAAAATCATGCGGTTTCACTTCATG	BamHI-NdeI
	Rev	CCCGCTCGAG-AGCTAATTGTGCTTGGTTTGCAGATAGGAGTT	XhoI
564c-(His/GST)(MC58)	Fwd	CGCGGATCCCATATG-GTTTCAGACGGCCTATACAACCAACATGGTGAAATT	BamHI-NdeI
	Rev	CCCGCTCGAG-GCGGTAACTGCCGCTTGCACTGAATCCGTAA	XhoI
564bc-(His/GST)(MC58)	Fwd	CGCGGATCCCATATG-ACGGGAGAAAATCATGCGGTTTCACTTCATG	BamHI-NdeI
	Rev	CCCGCTCGAG-GCGGTAACTGCCGCTTGCACTGAATCCGTAA	XhoI
564d-(His/GST)(MC58)	Fwd	CGCGGATCCCATATG-CAAAGCAAAGTCAAAGCAGACCATGCCTCCGTAA	BamHI-NdeI
	Rev	CCCGCTCGAG-TCTTTTCCTTTCAATTATAACTTTAGTAGGTTCAATTTTG GTCCCC	XhoI
564cd-(His/GST)(MC58)	Fwd	CGCGGATCCCATATG-GTTTCAGACGGCCTATACAACCAACATGGTGAAATT	BamHI-NdeI
	Rev	CCCGCTCGAG-TCTTTTCCTTTCAATTATAACTTTAGTAGGTTCAATTTTG GTCCCC	XhoI
570L	Fwd	GCGGCCATATG-ACCCGTTTGACCCGCG	NdeI
	Rev	GCGGCCTCGAG-TCAGCGGGCGTTCATTTCTT	XhoI
576-1L	Fwd	CGCGGATCCCATATG-AACACCATTTTCAAAATC	NdeI
	Rev	CCCGCTCGAG-TTAATTTACTTTTTTGATGTCG	XhoI

580L	Fwd	GCGGCCATATG-GATTCGCCCAAGGTCGG	NdeI
	Rev	GCGGCCTCGAG-CTACACTTCCCCCGAAGTGG	XhoI
583L	Fwd	CGCGGATCCCATATG-ATAGTTGACCAAAGCC	NdeI
	Rev	CCCGCTCGAG-TTATTTTTCCGATTTTTCGG	XhoI
593	Fwd	GCGGCCATATG-CTTGAACTGAACGGACT	NdeI
	Rev	GCGGCCTCGAG-TCAGCGGAAGCGGACGATT	XhoI
650 (His-GST) (MC58)	Fwd	CGCGGATCCCATATGTCCAAACTCAAAACCATCG	BamHI-NdeI
	Rev	CCCGCTCGAGGCTTCCAATCAGTTTGACC	XhoI
652	Fwd	GCGGCCATATG-AGCGCAATCGTTGATATTTTC	NdeI
	Rev	GCGGCCTCGAG-TTATTTGCCCAGTTGGTAGAATG	XhoI
664L	Fwd	GCGGCCATATG-GTGATACATCCGCACTACTTC	NdeI
	Rev	GCGGCCTCGAG-TCAAAATCGAGTTTTACACCA	XhoI
726	Fwd	GCGGCCATATG-ACCATCTATTTCAAAAACGG	NdeI
	Rev	GCGGCCTCGAG-TCAGCCGATGTTTAGCGTCCATT	XhoI
741-His(MC58)	Fwd	CGCGGATCCCATATG-AGCAGCGGAGGGGGTG	NdeI
	Rev	CCCGCTCGAG-TTGCTTGGCGGCAAGGC	XhoI
ΔG741-His(MC58)	Fwd	CGCGGATCCCATATG-GTCGCCGCCGACATCG	NdeI
	Rev	CCCGCTCGAG-TTGCTTGGCGGCAAGGC	XhoI
686-2-(His/GST) (MC58)	Fwd	CGCGGATCCCATATG-GGCGGTTCGGAAGGCG	BamHI-NdeI
	Rev	CCCGCTCGAG-TTGAACACTGATGTCTTTTCCGA	XhoI
719-(His/GST) (MC58)	Fwd	CGCGGATCCGCTAGC-AAACTGTCGTTGGTGTTAAC	BamHI-NheI
	Rev	CCCGCTCGAG-TTGACCCGCTCCACGG	XhoI
730-His (MC58)	Fwd	GCCGCCATATGGCGGACTTGGCGCAAGACCC	NdeI
	Rev	GCCGCCTCGAGATCTCCTAAACCTGTTTTAACAATGCCG	XhoI
730A-His (MC58)	Fwd	GCCGCCATATGGCGGACTTGGCGCAAGACCC	NdeI
	Rev	GCGGCCTCGAGCTCCATGCTGTTGCCCCAGC	XhoI
730B-His (MC58)	Fwd	GCCGCCATATGGCGGACTTGGCGCAAGACCC	NdeI
	Rev	GCGGCCTCGAGAAAATCCCCGCTAACCGCAG	XhoI
741-His (MC58)	Fwd	CGCGGATCCCATATG-AGCAGCGGAGGGGGTG	NdeI
	Rev	CCCGCTCGAG-TTGCTTGGCGGCAAGGC	XhoI
ΔG741-His (MC58)	Fwd	CGCGGATCCCATATG-GTCGCCGCCGACATCG	NdeI
	Rev	CCCGCTCGAG-TTGCTTGGCGGCAAGGC	XhoI
743 (His-GST)	Fwd	CGCGGATCCCATATGGACGGTGTTGTGCCTGTT	BamHI-NdeI
	Rev	CCCGCTCGAGCTTACGGATCAAATTGACG	XhoI
757 (His-GST) (MC58)	Fwd	CGCGGATCCCATATGGGCAGCCAATCTGAAGAA	BamHI-NdeI
	Rev	CCCGCTCGAGCTCAGCTTTTGCCGTCAA	XhoI
759-His/GST (MC58)	Fwd	CGCGGATCCGCTAGC-TACTCATCCATTGTCCGC	BamHI-NheI
	Rev	CCCGCTCGAG-CCAGTTGTAGCCTATTTTG	XhoI
759L (MC58)	Fwd	CGCGGATCCGCTAGC-ATGCGCTTCACACACAC	NheI
	Rev	CCCGCTCGAG-TTACCAGTTGTAGCCTATTT	XhoI
760-His	Fwd	GCCGCCATATGGCACAAACGGAAGGTTTGGAA	NdeI
	Rev	GCCGCCTCGAGAAAACTGTAACGCAGGTTTGCCGTC	XhoI
769-His (MC58)	Fwd	GCGGCCATATGGAAGAAACACCGCGCGAACCG	NdeI

	Rev	GCGGCCTCGAGGAACGTTTTATTAAACTCGAC	XhoI
907L	Fwd	GCGGCCATATG-AGAAAACCGACCGATACCCTA	NdeI
	Rev	GCGGCCTCGAG-TCAACGCCACTGCCAGCGGTTG	XhoI
911L	Fwd	CGCGGATCCCATATG-AAGAAGAACATATTGGAATTTTGGGTCGGACTG	NdeI
	Rev	CCCGCTCGAG-TTATTCGGCGGCTTTTTCCGCATTGCCG	XhoI
911LOmpA	Fwd	GGGAATTCCATATGAAAAAGACAGCTATCGCGATTGCA GTGGCACTGGCTGGTTTCGCTACCGTAGCGCAGGCCGC TAGC-GCTTTCCGCGTGGCCGGCGGTGC	NdeI-(NheI)
	Rev	CCCGCTCGAG-TTATTCGGCGGCTTTTTCCGCATTGCCG	XhoI
911LPelB	Fwd	CATGCCATGG-CTTTCCGCGTGGCCGGCGGTGC	NcoI
	Rev	CCCGCTCGAG-TTATTCGGCGGCTTTTTCCGCATTGCCG	XhoI
913-His/GST	Fwd	CGCGGATCCCATATG-TTTGCCGAAACCCGCC	BamHI-NdeI
(MC58)	Rev	CCCGCTCGAG-AGGTTGTGTTCCAGGTTG	XhoI
913L	Fwd	CGCGGATCCCATATG-AAAAAAACCGCCTATG	NdeI
(MC58)	Rev	CCCGCTCGAG-TTAAGGTTGTGTTCCAGG	XhoI
919L	Fwd	CGCGGATCCCATATG-AAAAAATACCTATTCCGC	NdeI
	Rev	CCCGCTCGAG-TTACGGGCGGTATTCGG	XhoI
919	Fwd	CGCGGATCCCATATG-CAAAGCAAGAGCATCCAAA	NdeI
	Rev	CCCGCTCGAG-TTACGGGCGGTATTCGG	XhoI
919L Orf4	Fwd	GGGAATTCCATATGAAAACCTTCTTCAAAACCCTTTCCG CCGCCGCGCTAGCGCTCATCCTCGCCGCC-TGCCAAAGCAAGAGCATC	NdeI-(NheI)
	Rev	CCCGCTCGAG-TTACGGGCGGTATTCGGGCTTCATACCG	XhoI
(919)-287fusion	Fwd	CGCGGATCCGTCGAC-TGTGGGGGCGGCGGTGGC	SalI
	Rev	CCCGCTCGAG-TCAATCCTGCTCTTTTTTGCC	XhoI
920-1L	Fwd	GCGGCCATATG-AAGAAAACATTGACACTGC	NdeI
	Rev	GCGGCCTCGAG-TTAATGGTGCGAATGACCGAT	XhoI
925-His/GST (MC58) GATE	Fwd	ggggacaagtttgtacaaaaaagcaggctTGCGGCAAGGATGCCGG	*attB1*
	Rev	ggggaccactttgtacaagaaagctgggtCTAAAGCAACAATGCCGG	*attB2*
926L	Fwd	CGCGGATCCCATATG-AAACACACCGTATCC	NdeI
	Rev	CCCGCTCGAG-TTATCTCGTGCGCGCC	XhoI
927-2-(His/GST)	Fwd	CGCGGATCCCATATG-AGCCCCGCGCCGATT	BamHI-NdeI
(MC58)	Rev	CCCGCTCGAG-TTTTTGTGCGGTCAGGCG	XhoI
932-His/GST (MC58) GATE	Fwd	ggggacaagtttgtacaaaaaagcaggctTGTTCGTTTGGGGGATTTAA ACCAAACCAAATC	*attB1*
935 (His-GST) (MC58)	For	CGCGGATCCCATATGGCGGATGCGCCCGCG	BamHI-NdeI
	Rev	CCCGCTCGAGAAACCGCCAATCCGCC	XhoI
	Rev	ggggaccactttgtacaagaaagctgggtTCATTTTGTTTTTCCTTCTTCT CGAGGCCATT	*attB2*
936-1L	Fwd	CGCGGATCCCATATG-AAACCCAAACCGCAC	NdeI
	Rev	CCCGCTCGAG-TCAGCGTTGGACGTAGT	XhoI
953L	Fwd	GGGAATTCCATATG-AAAAAAATCATCTTCGCCG	NdeI
	Rev	CCCGCTCGAG-TTATTGTTTGGCTGCCTCGAT	XhoI
953-fu	Fwd	GGGAATTCCATATG-GCCACCTACAAAGTGGACG	NdeI
	Rev	CGGGGATCC-TTGTTTGGCTGCCTCGATTTG	BamHI

954 (His-GST) (MC58)	Fwd	CGCGGATCCCATATGCAAGAACAATCGCAGAAAG	BamHI-NdeI
	Rev	CCCGCTCGAGTTTTTTCGGCAAATTGGCTT	XhoI
958-His/GST (MC58) GATE	Fwd	ggggacaagtttgtacaaaaaagcaggctGCCGATGCCGTTGCGG	*attB1*
	Rev	ggggaccactttgtacaagaaagctgggtTCAGGGTCGTTTGTTGCG	*attB2*
961L	Fwd	CGCGGATCCCATATG-AAACACTTTCCATCC	NdeI
	Rev	CCCGCTCGAG-TTACCACTCGTAATTGAC	XhoI
961	Fwd	CGCGGATCCCATATG-GCCACAAGCGACGAC	NdeI
	Rev	CCCGCTCGAG-TTACCACTCGTAATTGAC	XhoI
961 c (His/GST)	Fwd	CGCGGATCCCATATG-GCCACAAACGACG	BamHI-NdeI
	Rev	CCCGCTCGAG-ACCCACGTTGTAAGGTTG	XhoI
961 c-(His/GST) (MC58)	Fwd	CGCGGATCCCATATG-GCCACAAGCGACGACGA	BamHI-NdeI
	Rev	CCCGCTCGAG-ACCCACGTTGTAAGGTTG	XhoI
961 c-L	Fwd	CGCGGATCCCATATG-ATGAAACACTTTCCATCC	NdeI
	Rev	CCCGCTCGAG-TTAACCCACGTTGTAAGGT	XhoI
961 c-L (MC58)	Fwd	CGCGGATCCCATATG-ATGAAACACTTTCCATCC	NdeI
	Rev	CCCGCTCGAG-TTAACCCACGTTGTAAGGT	XhoI
961 d (His/GST)	Fwd	CGCGGATCCCATATG-GCCACAAACGACG	BamHI-NdeI
	Rev	CCCGCTCGAG-GTCTGACACTGTTTTATCC	XhoI
961 Δ1-L	Fwd	CGCGGATCCCATATG-ATGAAACACTTTCCATCC	NdeI
	Rev	CCCGCTCGAG-TTATGCTTTGGCGGCAAAG	XhoI
fu 961-...	Fwd	CGCGGATCCCATATG- GCCACAAACGACGAC	NdeI
	Rev	CGCGGATCC-CCACTCGTAATTGACGCC	BamHI
fu 961-... (MC58)	Fwd	CGCGGATCCCATATG-GCCACAAGCGACGAC	NdeI
	Rev	CGCGGATCC-CCACTCGTAATTGACGCC	BamHI
fu 961 c -...	Fwd	CGCGGATCCCATATG-GCCACAAACGACGAC	NdeI
	Rev	CGCGGATCC -ACCCACGTTGTAAGGTTG	BamHI
fu 961 c-L-...	Fwd	CGCGGATCCCATATG- ATGAAACACTTTCCATCC	NdeI
	Rev	CGCGGATCC -ACCCACGTTGTAAGGTTG	BamHI
fu (961)-741(MC58)-His	Fwd	CGCGGATCC -GGAGGGGGTGGTGTCG	BamHI
	Rev	CCCGCTCGAG-TTGCTTGGCGGCAAGGC	XhoI
fu (961)-983-His	Fwd	CGCGGATCC - GGCGGAGGCGGCACTT	BamHI
	Rev	CCCGCTCGAG-GAACCGGTAGCCTACG	XhoI
fu (961)- Orf46.1-His	Fwd	CGCGGATCCGGTGGTGGTGGT-TCAGATTTGGCAAACGATTC	BamHI
	Rev	CCCGCTCGAG-CGTATCATATTTCACGTGC	XhoI
fu (961 c-L)-741(MC58)	Fwd	CGCGGATCC -GGAGGGGGTGGTGTCG	BamHI
	Rev	CCCGCTCGAG-TTATTGCTTGGCGGCAAG	XhoI
fu (961c-L)-983	Fwd	CGCGGATCC - GGCGGAGGCGGCACTT	BamHI
	Rev	CCCGCTCGAG-TCAGAACCGGTAGCCTAC	XhoI
fu (961c-L)-Orf46.1	Fwd	CGCGGATCCGGTGGTGGTGGT-TCAGATTTGGCAAACGATTC	BamHI
	Rev	CCCGCTCGAG-TTACGTATCATATTTCACGTGC	XhoI
961-(His/GST)	Fwd	CGCGGATCCCATATG-GCCACAAGCGACGACG	BamHI-NdeI

(MC58)	Rev	CCCGCTCGAG-CCACTCGTAATTGACGCC	XhoI
961 Δ1-His	Fwd	CGCGGATCCCATATG-GCCACAAACGACGAC	NdeI
	Rev	CCCGCTCGAG-TGCTTTGGCGGCAAAGTT	XhoI
961a-(His/GST)	Fwd	CGCGGATCCCATATG-GCCACAAACGACGAC	BamHI-NdeI
	Rev	CCCGCTCGAG-TTTAGCAATATTATCTTTGTTCGTAGC	XhoI
961b-(His/GST)	Fwd	CGCGGATCCCATATG-AAAGCAAACCGTGCCGA	BamHI-NdeI
	Rev	CCCGCTCGAG-CCACTCGTAATTGACGCC	XhoI
961-His/GST GATE	Fwd	ggggacaagtttgtacaaaaaagcaggctGCAGCCACAAACGACGACG ATGTTAAAAAAGC	*attB1*
	Rev	ggggaccactttgtacaagaaagctgggtTTACCACTCGTAATTGACGC CGACATGGTAGG	*attB2*
982	Fwd	GCGGCCATATG-GCAGCAAAAGACGTACAGTT	NdeI
	Rev	GCGGCCTCGAG-TTACATCATGCCGCCCATACCA	XhoI
983-His (2996)	Fwd	CGCGGATCCGCTAGC-TTAGGCGGCGGCGGAG	NheI
	Rev	CCCGCTCGAG-GAACCGGTAGCCTACG	XhoI
ΔG983-His (2996)	Fwd	CCCCTAGCTAGC-ACTTCTGCGCCCGACTT	NheI
	Rev	CCCGCTCGAG-GAACCGGTAGCCTACG	XhoI
983-His	Fwd	CGCGGATCCGCTAGC-TTAGGCGGCGGCGGAG	NheI
	Rev	CCCGCTCGAG-GAACCGGTAGCCTACG	XhoI
ΔG983-His	Fwd	CGCGGATCCGCTAGC-ACTTCTGCGCCCGACTT	NheI
	Rev	CCCGCTCGAG-GAACCGGTAGCCTACG	XhoI
983L	Fwd	CGCGGATCCGCTAGC-CGAACGACCCCAACCTTCCCTACAAAAACTTTCAA	NheI
	Rev	CCCGCTCGAG-TCAGAACCGACGTGCCAAGCCGTTC	XhoI
987-His (MC58)	Fwd	GCCGCCATATGCCCCCACTGGAAGAACGGACG	NdeI
	Rev	GCCGCCTCGAGTAATAAACCTTCTATGGGCAGCAG	XhoI
989-(His/GST)	Fwd	CGCGGATCCCATATG-TCCGTCCACGCATCCG	BamHI-NdeI
(MC58)	Rev	CCCGCTCGAG-TTTGAATTTGTAGGTGTATTG	XhoI
989L	Fwd	CGCGGATCCCATATG-ACCCCTTCCGCACT	NdeI
(MC58)	Rev	CCCGCTCGAG-TTATTTGAATTTGTAGGTGTAT	XhoI
CrgA-His	Fwd	CGCGGATCCCATATG-AAAACCAATTCAGAAGAA	NdeI
(MC58)	Rev	CCCGCTCGAG-TCCACAGAGATTGTTTCC	XhoI
PilC1-ES	Fwd	GATGCCCGAAGGGCGGG	
(MC58)	Rev	GCCCAAGCTT-TCAGAAGAAGACTTCACGC	
PilC1-His	Fwd	CGCGGATCCCATATG-CAAACCCATAAATACGCTATT	NdeI
(MC58)	Rev	GCCCAAGCTT-GAAGAAGACTTCACGCCAG	HindIII
Δ1PilC1-His	Fwd	CGCGGATCCCATATG-GTCTTTTTCGACAATACCGA	NdeI
(MC58)	Rev	GCCCAAGCTT-	HindIII
PilC1L	Fwd	CGCGGATCCCATATG-AATAAAACTTTAAAAAGGCGG	NdeI
(MC58)	Rev	GCCCAAGCTT-TCAGAAGAAGACTTCACGC	HindIII
ΔGTbp2-His	Fwd	CGCGAATCCCATATG-TTCGATCTTGATTCTGTCGA	NdeI
(MC58)	Rev	CCCGCTCGAG-TCGCACAGGCTGTTGGCG	XhoI
Tbp2-His	Fwd	CGCGAATCCCATATG-TTGGGCGGAGGCGGCAG	NdeI
(MC58)	Rev	CCCGCTCGAG-TCGCACAGGCTGTTGGCG	XhoI
Tbp2-His(MC58)	Fwd	CGCGAATCCCATATG-TTGGGCGGAGGCGGCAG	NdeI
	Rev	CCCGCTCGAG-TCGCACAGGCTGTTGGCG	XhoI

NMB0109-(His/GST) (MC58)	Fwd	CGCGGATCCCATATG-GCAAATTTGGAGGTGCGC	BamHI-NdeI
	Rev	CCCGCTCGAG-TTCGGAGCGGTTGAAGC	XhoI
NMB0109L (MC58)	Fwd	CGCGGATCCCATATG-CAACGTCGTATTATAACCC	NdeI
	Rev	CCCGCTCGAG-TTATTCGGAGCGGTTGAAG	XhoI
NMB0207-(His/GST) (MC58)	Fwd	CGCGGATCCCATATG-GGCATCAAAGTCGCCATCAACGGCTAC	BamHI-NdeI
	Rev	CCCGCTCGAG-TTTGAGCGGGCGCACTTCAAGTCCG	XhoI
NMB0462-(His/GST) (MC58)	Fwd	CGCGGATCCCATATG-GGCGGCAGCGAAAAAAAC	BamHI-NdeI
	Rev	CCCGCTCGAG-GTTGGTGCCGACTTTGAT	XhoI
NMB0623-(His/GST) (MC58)	Fwd	CGCGGATCCCATATG-GGCGGCGGAAGCGATA	BamHI-NdeI
	Rev	CCCGCTCGAG-TTTGCCCGCTTTGAGCC	XhoI
NMB0625 (His-GST)(MC58)	Fwd	CGCGGATCCCATATGGGCAAATCCGAAAATACG	BamHI-NdeI
	Rev	CCCGCTCGAGCATCCCGTACTGTTTCG	XhoI
NMB0634 (His/GST)(MC58)	Fwd	ggggacaagtttgtacaaaaaagcaggctCCGACATTACCGTGTACAAC GGCCAACAAAGAA	*attB1*
	Rev	ggggaccactttgtacaagaaagctgggtCTTATTTCATACCGGCTTGCT CAAGCAGCCGG	*attB2*
NMB0776-His/GST (MC58) GATE	Fwd	ggggacaagtttgtacaaaaaagcaggctGATACGGTGTTTTCCTGTAA AACGGACAACAA	*attB1*
	Rev	ggggaccactttgtacaagaaagctgggtCTAGGAAAAATCGTCATCGT TGAAATTCGCC	*attB2*
NMB1115-His/GST (MC58) GATE	Fwd	ggggacaagtttgtacaaaaaagcaggctATGCACCCCATCGAAACC	*attB1*
	Rev	ggggaccactttgtacaagaaagctgggtCTAGTCTTGCAGTGCCTC	*attB2*
NMB1343-(His/GST) (MC58)	Fwd	CGCGGATCCCATATG-GGAAATTTCTTATATAGAGGCATTAG	BamHI-NdeI
	Rev	CCCGCTCGAG-GTTAATTTCTATCAACTCTTTAGCAATAAT	XhoI
NMB1369 (His-GST (MC58)	Fwd	CGCGGATCCCATATGGCCTGCCAAGACGACA	BamHI-NdeI
	Rev	CCCGCTCGAGCCGCCTCCTGCCGAAA	XhoI
NMB1551 (His-GST)(MC58)	Fwd	CGCGGATCCCATATGGCAGAGATCTGTTTGATAA	BamHI-NdeI
	Rev	CCCGCTCGAGCGGTTTTCCGCCCAATG	XhoI
NMB1899 (His-GST) (MC58)	Fwd	CGCGGATCCCATATGCAGCCGGATACGGTC	BamHI-NdeI
	Rev	CCCGCTCGAGAATCACTTCCAACACAAAAT	XhoI
NMB2050-(His/GST) (MC58)	Fwd	CGCGGATCCCATATG-TGGTTGCTGATGAAGGGC	BamHI-NdeI
	Rev	CCCGCTCGAG-GACTGCTTCATCTTCTGC	XhoI
NMB2050L (MC58)	Fwd	CGCGGATCCCATATG-GAACTGATGACTGTTTTGC	NdeI
	Rev	CCCGCTCGAG-TCAGACTGCTTCATCTTCT	XhoI
NMB2159-(His/GST) (MC58)	Fwd	CGCGGATCCCATATG-AGCATTAAAGTAGCGATTAACGGTTTCGGC	BamHI-NdeI
	Rev	CCCGCTCGAG-GATTTTGCCTGCGAAGTATTCCAAAGTGCG	XhoI
fu-ΔG287...-His	Fwd	CGCGGATCCGCTAGC-CCCGATGTTAAATCGGC	NheI

	Rev	CGGGGATCC-ATCCTGCTCTTTTTTGCCGG	BamHI
fu-(ΔG287)-919-His	Fwd	CGCGGATCCGGTGGTGGTGGT-CAAAGCAAGAGCATCCAAACC	BamHI
	Rev	CCCAAGCTT-TTCGGGCGGTATTCGGGCTTC	HindIII
fu-(ΔG287)-953-His	Fwd	CGCGGATCCGGTGGTGGTGGT-GCCACCTACAAAGTGGAC	BamHI
	Rev	GCCCAAGCTT-TTGTTTGGCTGCCTCGAT	HindIII
fu-(ΔG287)-961-His	Fwd	CGCGGATCCGGTGGTGGTGGT-ACAAGCGACGACG	BamHI
	Rev	GCCCAAGCTT-CCACTCGTAATTGACGCC	HindIII
fu-(ΔG287)-Orf46.1-His	Fwd	CGCGGATCCGGTGGTGGTGGT-TCAGATTTGGCAAACGATTC	BamHI
	Rev	CCCAAGCTT-CGTATCATATTTCACGTGC	HindIII
fu-(ΔG287-919)-Orf46.1-His	Fwd	CCCAAGCTTGGTGGTGGTGGTGGT-TCAGATTTGGCAAACGATTC	HindIII
	Rev	CCCGCTCGAG-CGTATCATATTTCACGTGC	XhoI
fu-(ΔG287-Orf46.1)-919-His	Fwd	CCCAAGCTTGGTGGTGGTGGTGGT-CAAAGCAAGAGCATCCAAACC	HindIII
	Rev	CCCGCTCGAG-CGGGCGGTATTCGGGCTT	XhoI
fu ΔG287(394.98)-...	Fwd	CGCGGATCCGCTAGC-CCCGATGTTAAATCGGC	NheI
	Rev	CGGGGATCC-ATCCTGCTCTTTTTTGCCGG	BamHI
fu Orf1-(Orf46.1)-His	Fwd	CGCGGATCCGCTAGC-GGACACACTTATTTCGGCATC	NheI
	Rev	CGCGGATCC-CCAGCGGTAGCCTAATTTGAT	
fu (Orf1)-Orf46.1-His	Fwd	CGCGGATCCGGTGGTGGTGGT-TCAGATTTGGCAAACGATTC	BamHI
	Rev	CCCAAGCTT-CGTATCATATTTCACGTGC	HindIII
fu (919)-Orf46.1-His	Fwd1	GCGGCGTCGACGGTGGCGGAGGCACTGGATCCTCAG	SalI
	Fwd2	GGAGGCACTGGATCCTCAGATTTGGCAAACGATTC	
	Rev	CCCGCTCGAG-CGTATCATATTTCACGTGC	XhoI
Fu orf46-....	Fwd	GGAATTCCATATGTCAGATTTGGCAAACGATTC	NdeI
	Rev	CGCGGATCCCGTATCATATTTCACGTGC	BamHI
Fu (orf46)-287-His	Fwd	CGGGGATCCGGGGGCGGCGGTGGCG	BamHI
	Rev	CCCAAGCTTATCCTGCTCTTTTTTGCCGGC	HindIII
Fu (orf46)-919-His	Fwd	CGCGGATCCGGTGGTGGTGGTCAAAGCAAGAGCATCCAAACC	BamHI
	Rev	CCCAAGCTTCGGGCGGTATTCGGGCTTC	HindIII
Fu (orf46-919)-287-His	Fwd	CCCCAAGCTTGGGGGCGGCGGTGGCG	HindIII
	Rev	CCCGCTCGAGATCCTGCTCTTTTTTGCCGGC	XhoI
Fu (orf46-287)-919-His	Fwd	CCCAAGCTTGGTGGTGGTGGTGGTCAAAGCAAGAGCATCCAAACC	HindIII
	Rev	CCCGCTCGAGCGGGCGGTATTCGGGCTT	XhoI
(ΔG741)-961c-His	Fwd1	GGAGGCACTGGATCCGCAGCCACAAACGACGACGA	XhoI
	Fwd2	GCGGCCTCGAG-GGTGGCGGAGGCACTGGATCCGCAG	
	Rev	CCCGCTCGAG-ACCCAGCTTGTAAGGTTG	XhoI
(ΔG741)-961-His	Fwd1	GGAGGCACTGGATCCGCAGCCACAAACGACGACGA	XhoI
	Fwd2	GCGGCCTCGAG-GGTGGCGGAGGCACTGGATCCGCAG	
	Rev	CCCGCTCGAG-CCACTCGTAATTGACGCC	XhoI

(ΔG741)-983-His	Fwd	GCGGCCTCGAG-GGATCCGGCGGAGGCGGCACTTCTGCG	XhoI
	Rev	CCCGCTCGAG-GAACCGGTAGCCTACG	XhoI
(ΔG741)-orf46.1-His	Fwd1	GGAGGCACTGGATCCTCAGATTTGGCAAACGATTC	SalI
	Fwd2	GCGGCGTCGACGGTGGCGGAGGCACTGGATCCTCAGA	
	Rev	CCCGCTCGAG-CGTATCATATTTCACGTGC	XhoI
(ΔG983)-741(MC58) -His	Fwd	GCGGCCTCGAG-GGATCCGGAGGGGGTGGTGTCGCC	XhoI
	Rev	CCCGCTCGAG-TTGCTTGGCGGCAAG	XhoI
(ΔG983)-961c-His	Fwd1	GGAGGCACTGGATCCGCAGCCACAAACGACGACGA	XhoI
	Fwd2	GCGGCCTCGAG-GGTGGCGGAGGCACTGGATCCGCAG	
	Rev	CCCGCTCGAG-ACCCAGCTTGTAAGGTTG	XhoI
(ΔG983)-961-His	Fwd1	GGAGGCACTGGATCCGCAGCCACAAACGACGACGA	XhoI
	Fwd2	GCGGCCTCGAG-GGTGGCGGAGGCACTGGATCCGCAG	
	Rev	CCCGCTCGAG-CCACTCGTAATTGACGCC	XhoI
(ΔG983)- Orf46.1-His	Fwd1	GGAGGCACTGGATCCTCAGATTTGGCAAACGATTC	SalI
	Fwd2	GCGGCGTCGACGGTGGCGGAGGCACTGGATCCTCAGA	
	Rev	CCCGCTCGAG-CGTATCATATTTCACGTGC	XhoI

* This primer was used as a Reverse primer for all the C terminal fusions of 287 to the His-tag.

§ Forward primers used in combination with the 287-His Reverse primer.

NB – All PCR reactions use strain 2996 unless otherwise specified (*e.g.* strain MC58)

In all constructs starting with an ATG not followed by a unique *Nhe*I site, the ATG codon is part of the *Nde*I site used for cloning. The constructs made using *Nhe*I as a cloning site at the 5' end (*e.g.* all those containing 287 at the N-terminus) have two additional codons (GCT AGC) fused to the coding sequence of the antigen.

Preparation of chromosomal DNA templates

N.meningitidis strains 2996, MC58, 394.98, 1000 and BZ232 (and others) were grown to exponential phase in 100ml of GC medium, harvested by centrifugation, and resuspended in 5ml buffer (20% w/v sucrose, 50mM Tris-HCl, 50mM EDTA, pH8). After 10 minutes incubation on ice, the bacteria were lysed by adding 10ml of lysis solution (50mM NaCl, 1% Na-Sarkosyl, 50μg/ml Proteinase K), and the suspension incubated at 37°C for 2 hours. Two phenol extractions (equilibrated to pH 8) and one $CHCl_3$/isoamylalcohol (24:1) extraction were performed. DNA was precipitated by addition of 0.3M sodium acetate and 2 volumes of ethanol, and collected by centrifugation. The pellet was washed once with 70%(v/v) ethanol and redissolved in 4.0ml TE buffer (10mM Tris-HCl, 1mM EDTA, pH 8.0). The DNA concentration was measured by reading OD_{260}.

PCR Amplification

The standard PCR protocol was as follows: 200ng of genomic DNA from 2996, MC581000, or BZ232 strains or 10ng of plasmid DNA preparation of recombinant clones were used as template in the presence of 40μM of each oligonucletide primer, 400-800 μM dNTPs solution, 1x PCR buffer (including 1.5mM $MgCl_2$), 2.5 units *TaqI* DNA polymerase (using Perkin-Elmer AmpliTaQ, Boerhingher Mannheim Expand™ Long Template).

After a preliminary 3 minute incubation of the whole mix at 95°C, each sample underwent a two-step amplification: the first 5 cycles were performed using the hybridisation temperature that excluded the restriction enzyme tail of the primer (T_{m1}). This was followed by 30 cycles according to the hybridisation temperature calculated for the whole length oligos (T_{m2}). Elongation times, performed at 68°C or 72°C, varied according to the length of the Orf to be amplified. In the case of Orf1 the elongation time, starting from 3 minutes, was increased by 15 seconds each cycle. The cycles were completed with a 10 minute extension step at 72°C.

The amplified DNA was either loaded directly on a 1% agarose gel. The DNA fragment corresponding to the band of correct size was purified from the gel using the Qiagen Gel Extraction Kit, following the manufacturer's protocol.

Digestion of PCR fragments and of the cloning vectors

The purified DNA corresponding to the amplified fragment was digested with the appropriate restriction enzymes for cloning into pET-21b+, pET22b+ or pET-24b+. Digested fragments were purified using the QIAquick PCR purification kit (following the manufacturer's instructions) and eluted with either H_2O or 10mM Tris, pH 8.5. Plasmid vectors were digested with the appropriate restriction enzymes, loaded onto a 1.0% agarose gel and the band corresponding to the digested vector purified using the Qiagen QIAquick Gel Extraction Kit.

Cloning

The fragments corresponding to each gene, previously digested and purified, were ligated into pET21b+, pET22b+ or pET-24b+. A molar ratio of 3:1 fragment/vector was used with T4 DNA ligase in the ligation buffer supplied by the manufacturer.

Recombinant plasmid was transformed into competent *E.coli* DH5 or HB101 by incubating the ligase reaction solution and bacteria for 40 minutes on ice, then at 37°C for 3 minutes.

This was followed by the addition of 800µl LB broth and incubation at 37°C for 20 minutes. The cells were centrifuged at maximum speed in an Eppendorf microfuge, resuspended in approximately 200µl of the supernatant and plated onto LB ampicillin (100mg/ml) agar.

Screening for recombinant clones was performed by growing randomly selected colonies overnight at 37°C in 4.0ml of LB broth + 100µg/ml ampicillin. Cells were pelleted and plasmid DNA extracted using the Qiagen QIAprep Spin Miniprep Kit, following the manufacturer's instructions. Approximately 1µg of each individual miniprep was digested with the appropriate restriction enzymes and the digest loaded onto a 1-1.5% agarose gel (depending on the expected insert size), in parallel with the molecular weight marker (1kb DNA Ladder, GIBCO). Positive clones were selected on the basis of the size of insert.

Expression

After cloning each gene into the expression vector, recombinant plasmids were transformed into *E.coli* strains suitable for expression of the recombinant protein. 1µl of each construct was used to transform *E.coli* BL21-DE3 as described above. Single recombinant colonies were inoculated into 2ml LB+Amp (100µg/ml), incubated at 37°C overnight, then diluted 1:30 in 20ml of LB+Amp (100µg/ml) in 100ml flasks, to give an OD_{600} between 0.1 and 0.2. The flasks were incubated at 30°C or at 37°C in a gyratory water bath shaker until OD_{600} indicated exponential growth suitable for induction of expression (0.4-0.8 OD). Protein expression was induced by addition of 1.0mM IPTG. After 3 hours incubation at 30°C or 37°C the OD_{600} was measured and expression examined. 1.0ml of each sample was centrifuged in a microfuge, the pellet resuspended in PBS and analysed by SDS-PAGE and Coomassie Blue staining.

Gateway cloning and expression

Sequences labelled GATE were cloned and expressed using the GATEWAY Cloning Technology (GIBCO-BRL). Recombinational cloning (RC) is based on the recombination reactions that mediate the integration and excision of phage into and from the *E.coli* genome, respectively. The integration involves recombination of the *attP* site of the phage DNA within the *attB* site located in the bacterial genome (BP reaction) and generates an integrated phage genome flanked by *attL* and *attR* sites. The excision recombines *attL* and *attR* sites back to *attP* and *attB* sites (LR reaction). The integration reaction requires two enzymes [the phage protein Integrase (Int) and the bacterial protein integration host factor (IHF)] (BP clonase). The

excision reaction requires Int, IHF, and an additional phage enzyme, Excisionase (Xis) (LR clonase). Artificial derivatives of the 25-bp bacterial *attB* recombination site, referred to as B1 and B2, were added to the 5' end of the primers used in PCR reactions to amplify Neisserial ORFs. The resulting products were BP cloned into a "Donor vector" containing complementary derivatives of the phage *attP* recombination site (P1 and P2) using BP clonase. The resulting "Entry clones" contain ORFs flanked by derivatives of the *attL* site (L1 and L2) and were subcloned into expression "destination vectors" which contain derivatives of the *attL*-compatible *attR* sites (R1 and R2) using LR clonase. This resulted in "expression clones" in which ORFs are flanked by B1 and B2 and fused in frame to the GST or His N terminal tags.

The *E. coli* strain used for GATEWAY expression is BL21-SI. Cells of this strain are induced for expression of the T7 RNA polymerase by growth in medium containing salt (0.3 M NaCl).

Note that this system gives N-terminus His tags.

Preparation of membrane proteins.

Fractions composed principally of either inner, outer or total membrane were isolated in order to obtain recombinant proteins expressed with membrane-localisation leader sequences. The method for preparation of membrane fractions, enriched for recombinant proteins, was adapted from Filip *et. al.* [*J.Bact.* (1973) 115:717-722] and Davies *et. al.* [*J.Immunol.Meth.* (1990) 143:215-225]. Single colonies harbouring the plasmid of interest were grown overnight at 37°C in 20 ml of LB/Amp (100 µg/ml) liquid culture. Bacteria were diluted 1:30 in 1.0 L of fresh medium and grown at either 30°C or 37°C until the OD_{550} reached 0.6-0.8. Expression of recombinant protein was induced with IPTG at a final concentration of 1.0 mM. After incubation for 3 hours, bacteria were harvested by centrifugation at 8000*g* for 15 minutes at 4°C and resuspended in 20 ml of 20 mM Tris-HCl (pH 7.5) and complete protease inhibitors (Boehringer-Mannheim). All subsequent procedures were performed at 4°C or on ice.

Cells were disrupted by sonication using a Branson Sonifier 450 and centrifuged at 5000*g* for 20 min to sediment unbroken cells and inclusion bodies. The supernatant, containing membranes and cellular debris, was centrifuged at 50000*g* (Beckman Ti50, 29000rpm) for 75 min, washed with 20 mM Bis-tris propane (pH 6.5), 1.0 M NaCl, 10% (v/v) glycerol and sedimented again at 50000*g* for 75 minutes. The pellet was resuspended in 20mM Tris-HCl (pH 7.5), 2.0% (v/v) Sarkosyl, complete protease inhibitor (1.0 mM EDTA, final

concentration) and incubated for 20 minutes to dissolve inner membrane. Cellular debris was pelleted by centrifugation at 5000*g* for 10 min and the supernatant centrifuged at 75000*g* for 75 minutes (Beckman Ti50, 33000rpm). Proteins 008L and 519L were found in the supernatant suggesting inner membrane localisation. For these proteins both inner and total membrane fractions (washed with NaCl as above) were used to immunise mice. Outer membrane vesicles obtained from the 75000*g* pellet were washed with 20 mM Tris-HCl (pH 7.5) and centrifuged at 75000*g* for 75 minutes or overnight. The OMV was finally resuspended in 500 µl of 20 mM Tris-HCl (pH 7.5), 10% v/v glycerol. Orf1L and Orf40L were both localised and enriched in the outer membrane fraction which was used to immunise mice. Protein concentration was estimated by standard Bradford Assay (Bio-Rad), while protein concentration of inner membrane fraction was determined with the DC protein assay (Bio-Rad). Various fractions from the isolation procedure were assayed by SDS-PAGE.

Purification of His-tagged proteins

Various forms of 287 were cloned from strains 2996 and MC58. They were constructed with a C-terminus His-tagged fusion and included a mature form (aa 18-427), constructs with deletions (Δ1, Δ 2, Δ3 and Δ4) and clones composed of either B or C domains. For each clone purified as a His-fusion, a single colony was streaked and grown overnight at 37°C on a LB/Amp (100 µg/ml) agar plate. An isolated colony from this plate was inoculated into 20ml of LB/Amp (100 µg/ml) liquid medium and grown overnight at 37°C with shaking. The overnight culture was diluted 1:30 into 1.0 L LB/Amp (100 µg/ml) liquid medium and allowed to grow at the optimal temperature (30 or 37°C) until the OD_{550} reached 0.6-0.8. Expression of recombinant protein was induced by addition of IPTG (final concentration 1.0mM) and the culture incubated for a further 3 hours. Bacteria were harvested by centrifugation at 8000*g* for 15 min at 4°C. The bacterial pellet was resuspended in 7.5 ml of either (i) cold buffer A (300 mM NaCl, 50 mM phosphate buffer, 10 mM imidazole, pH 8.0) for soluble proteins or (ii) buffer B (10mM Tris-HCl, 100 mM phosphate buffer, pH 8.8 and, optionally, 8M urea) for insoluble proteins. Proteins purified in a soluble form included 287-His, Δ1, Δ2, Δ3 and Δ4287-His, Δ4287MC58-His, 287c-His and 287cMC58-His. Protein 287bMC58-His was insoluble and purified accordingly. Cells were disrupted by sonication on ice four times for 30 sec at 40W using a Branson sonifier 450 and centrifuged at 13000xg for 30 min at 4°C. For insoluble proteins, pellets were resuspended in 2.0 ml buffer C (6 M guanidine hydrochloride, 100 mM phosphate buffer, 10 mM Tris- HCl, pH 7.5

and treated with 10 passes of a Dounce homogenizer. The homogenate was centrifuged at 13000*g* for 30 min and the supernatant retained. Supernatants for both soluble and insoluble preparations were mixed with 150µl Ni^{2+}-resin (previously equilibrated with either buffer A or buffer B, as appropriate) and incubated at room temperature with gentle agitation for 30 min. The resin was Chelating Sepharose Fast Flow (Pharmacia), prepared according to the manufacturer's protocol. The batch-wise preparation was centrifuged at 700*g* for 5 min at 4°C and the supernatant discarded. The resin was washed twice (batch-wise) with 10ml buffer A or B for 10 min, resuspended in 1.0 ml buffer A or B and loaded onto a disposable column. The resin continued to be washed with either (i) buffer A at 4°C or (ii) buffer B at room temperature, until the OD_{280} of the flow-through reached 0.02-0.01. The resin was further washed with either (i) cold buffer C (300mM NaCl, 50mM phosphate buffer, 20mM imidazole, pH 8.0) or (ii) buffer D (10mM Tris-HCl, 100mM phosphate buffer, pH 6.3 and, optionally, 8M urea) until OD_{280} of the flow-through reached 0.02-0.01. The His-fusion protein was eluted by addition of 700µl of either (i) cold elution buffer A (300 mM NaCl, 50mM phosphate buffer, 250 mM imidazole, pH 8.0) or (ii) elution buffer B (10 mM Tris-HCl, 100 mM phosphate buffer, pH 4.5 and, optionally, 8M urea) and fractions collected until the OD_{280} indicated all the recombinant protein was obtained. 20µl aliquots of each elution fraction were analysed by SDS-PAGE. Protein concentrations were estimated using the Bradford assay.

Renaturation of denatured His-fusion proteins.

Denaturation was required to solubilize 287bMC8, so a renaturation step was employed prior to immunisation. Glycerol was added to the denatured fractions obtained above to give a final concentration of 10% v/v. The proteins were diluted to 200 µg/ml using dialysis buffer I (10% v/v glycerol, 0.5M arginine, 50 mM phosphate buffer, 5.0 mM reduced glutathione, 0.5 mM oxidised glutathione, 2.0M urea, pH 8.8) and dialysed against the same buffer for 12-14 hours at 4°C. Further dialysis was performed with buffer II (10% v/v glycerol, 0.5M arginine, 50mM phosphate buffer, 5.0mM reduced glutathione, 0.5mM oxidised glutathione, pH 8.8) for 12-14 hours at 4°C. Protein concentration was estimated using the formula:

$$Protein\ (mg/ml) = (1.55 \times OD_{280}) - (0.76 \times OD_{260})$$

Amino acid sequence analysis.

Automated sequence analysis of the NH_2-terminus of proteins was performed on a Beckman sequencer (LF 3000) equipped with an on-line phenylthiohydantoin-amino acid analyser (System Gold) according to the manufacturer's recommendations.

Immunization

Balb/C mice were immunized with antigens on days 0, 21 and 35 and sera analyzed at day 49.

Sera analysis – ELISA

The acapsulated MenB M7 and the capsulated strains were plated on chocolate agar plates and incubated overnight at 37°C with 5% CO_2. Bacterial colonies were collected from the agar plates using a sterile dracon swab and inoculated into Mueller-Hinton Broth (Difco) containing 0.25% glucose. Bacterial growth was monitored every 30 minutes by following OD_{620}. The bacteria were let to grow until the OD reached the value of 0.4-0.5. The culture was centrifuged for 10 minutes at 4000rpm. The supernatant was discarded and bacteria were washed twice with PBS, resuspended in PBS containing 0.025% formaldehyde, and incubated for 1 hour at 37°C and then overnight at 4°C with stirring. 100µl bacterial cells were added to each well of a 96 well Greiner plate and incubated overnight at 4°C. The wells were then washed three times with PBT washing buffer (0.1% Tween-20 in PBS). 200µl of saturation buffer (2.7% polyvinylpyrrolidone 10 in water) was added to each well and the plates incubated for 2 hours at 37°C. Wells were washed three times with PBT. 200µl of diluted sera (Dilution buffer: 1% BSA, 0.1% Tween-20, 0.1% NaN_3 in PBS) were added to each well and the plates incubated for 2 hours at 37°C. Wells were washed three times with PBT. 100µl of HRP-conjugated rabbit anti-mouse (Dako) serum diluted 1:2000 in dilution buffer were added to each well and the plates were incubated for 90 minutes at 37°C. Wells were washed three times with PBT buffer. 100µl of substrate buffer for HRP (25ml of citrate buffer pH5, 10mg of O-phenildiamine and 10µl of H_2O_2) were added to each well and the plates were left at room temperature for 20 minutes. 100µl 12.5% H_2SO_4 was added to each well and OD_{490} was followed. The ELISA titers were calculated abitrarely as the dilution of sera which gave an OD_{490} value of 0.4 above the level of preimmune sera. The ELISA was considered positive when the dilution of sera with OD_{490} of 0.4 was higher than 1:400.

Sera analysis – FACS Scan bacteria binding assay

The acapsulated MenB M7 strain was plated on chocolate agar plates and incubated overnight at 37°C with 5% CO_2. Bacterial colonies were collected from the agar plates using

a sterile dracon swab and inoculated into 4 tubes containing 8ml each Mueller-Hinton Broth (Difco) containing 0.25% glucose. Bacterial growth was monitored every 30 minutes by following OD_{620}. The bacteria were let to grow until the OD reached the value of 0.35-0.5. The culture was centrifuged for 10 minutes at 4000rpm. The supernatant was discarded and the pellet was resuspended in blocking buffer (1% BSA in PBS, 0.4% NaN_3) and centrifuged for 5 minutes at 4000rpm. Cells were resuspended in blocking buffer to reach OD_{620} of 0.05. 100µl bacterial cells were added to each well of a Costar 96 well plate. 100µl of diluted (1:100, 1:200, 1:400) sera (in blocking buffer) were added to each well and plates incubated for 2 hours at 4°C. Cells were centrifuged for 5 minutes at 4000rpm, the supernatant aspirated and cells washed by addition of 200µl/well of blocking buffer in each well. 100µl of R-Phicoerytrin conjugated $F(ab)_2$ goat anti-mouse, diluted 1:100, was added to each well and plates incubated for 1 hour at 4°C. Cells were spun down by centrifugation at 4000rpm for 5 minutes and washed by addition of 200µl/well of blocking buffer. The supernatant was aspirated and cells resuspended in 200µl/well of PBS, 0.25% formaldehyde. Samples were transferred to FACScan tubes and read. The condition for FACScan (Laser Power 15mW) setting were: FL2 on; FSC-H threshold:92; FSC PMT Voltage: E 01; SSC PMT: 474; Amp. Gains 6.1; FL-2 PMT: 586; compensation values: 0.

Sera analysis – bactericidal assay

N. meningitidis strain 2996 was grown overnight at 37°C on chocolate agar plates (starting from a frozen stock) with 5% CO_2. Colonies were collected and used to inoculate 7ml Mueller-Hinton broth, containing 0.25% glucose to reach an OD_{620} of 0.05-0.08. The culture was incubated for approximately 1.5 hours at 37 degrees with shacking until the OD_{620} reached the value of 0.23-0.24. Bacteria were diluted in 50mM Phosphate buffer pH 7.2 containing 10mM $MgCl_2$, 10mM $CaCl_2$ and 0.5% (w/v) BSA (assay buffer) at the working dilution of 10^5 CFU/ml. The total volume of the final reaction mixture was 50 µl with 25 µl of serial two fold dilution of test serum, 12.5 µl of bacteria at the working dilution, 12.5 µl of baby rabbit complement (final concentration 25%).

Controls included bacteria incubated with complement serum, immune sera incubated with bacteria and with complement inactivated by heating at 56°C for 30'. Immediately after the addition of the baby rabbit complement, 10µl of the controls were plated on Mueller-Hinton agar plates using the tilt method (time 0). The 96-wells plate was incubated for 1 hour at 37°C with rotation. 7µl of each sample were plated on Mueller-Hinton agar plates as spots, whereas 10µl of the controls were plated on Mueller-Hinton agar plates using the tilt method

(time 1). Agar plates were incubated for 18 hours at 37 degrees and the colonies corresponding to time 0 and time 1 were counted.

Sera analysis – western blots

Purified proteins (500ng/lane), outer membrane vesicles (5µg) and total cell extracts (25µg) derived from MenB strain 2996 were loaded onto a 12% SDS-polyacrylamide gel and transferred to a nitrocellulose membrane. The transfer was performed for 2 hours at 150mA at 4°C, using transfer buffer (0.3% Tris base, 1.44% glycine, 20% (v/v) methanol). The membrane was saturated by overnight incubation at 4°C in saturation buffer (10% skimmed milk, 0.1% Triton X100 in PBS). The membrane was washed twice with washing buffer (3% skimmed milk, 0.1% Triton X100 in PBS) and incubated for 2 hours at 37°C with mice sera diluted 1:200 in washing buffer. The membrane was washed twice and incubated for 90 minutes with a 1:2000 dilution of horseradish peroxidase labelled anti-mouse Ig. The membrane was washed twice with 0.1% Triton X100 in PBS and developed with the Opti-4CN Substrate Kit (Bio-Rad). The reaction was stopped by adding water.

The OMVs were prepared as follows: *N. meningitidis* strain 2996 was grown overnight at 37 degrees with 5% CO_2 on 5 GC plates, harvested with a loop and resuspended in 10 ml of 20mM Tris-HCl pH 7.5, 2 mM EDTA. Heat inactivation was performed at 56°C for 45 minutes and the bacteria disrupted by sonication for 5 minutes on ice (50% duty cycle, 50% output , Branson sonifier 3 mm microtip). Unbroken cells were removed by centrifugation at 5000*g* for 10 minutes, the supernatant containing the total cell envelope fraction recovered and further centrifuged overnight at 50000*g* at the temperature of 4°C . The pellet containing the membranes was resuspended in 2% sarkosyl, 20mM Tris-HCl pH 7.5, 2 mM EDTA and incubated at room temperature for 20 minutes to solubilise the inner membranes. The suspension was centrifuged at 10000*g* for 10 minutes to remove aggregates, the supernatant was further centrifuged at 50000*g* for 3 hours. The pellet, containing the outer membranes was washed in PBS and resuspended in the same buffer. Protein concentration was measured by the D.C. Bio-Rad Protein assay (Modified Lowry method), using BSA as a standard.

Total cell extracts were prepared as follows: *N. meningitidis* strain 2996 was grown overnight on a GC plate, harvested with a loop and resuspended in 1ml of 20mM Tris-HCl. Heat inactivation was performed at 56°C for 30 minutes.

961 domain studies

Cellular fractions preparation Total lysate, periplasm, supernatant and OMV of *E.coli* clones expressing different domains of 961 were prepared using bacteria from over-night cultures or

after 3 hours induction with IPTG. Briefly, the periplasm were obtained suspending bacteria in saccarose 25% and Tris 50mM (pH 8) with polimixine 100µg/ml. After 1hr at room temperature bacteria were centrifuged at 13000rpm for 15 min and the supernatant were collected. The culture supernatant were filtered with 0.2µm and precipitated with TCA 50% in ice for two hours. After centrifugation (30 min at 13000 rp) pellets were rinsed twice with ethanol 70% and suspended in PBS. The OMV preparation was performed as previously described. Each cellular fraction were analyzed in SDS-PAGE or in Western Blot using the polyclonal anti-serum raised against GST-961.

Adhesion assay Chang epithelial cells (Wong-Kilbourne derivative, clone 1-5c-4, human conjunctiva) were maintained in DMEM (Gibco) supplemented with 10% heat-inactivated FCS, 15mM L-glutamine and antibiotics.

For the adherence assay, sub-confluent culture of Chang epithelial cells were rinsed with PBS and treated with trypsin-EDTA (Gibco), to release them from the plastic support. The cells were then suspended in PBS, counted and dilute in PBS to $5x10^5$ cells/ml.

Bacteria from over-night cultures or after induction with IPTG, were pelleted and washed twice with PBS by centrifuging at 13000 for 5 min. Approximately $2\text{-}3x10^8$ (cfu) were incubated with 0.5 mg/ml FITC (Sigma) in 1ml buffer containing 50mM $NaHCO_3$ and 100mM NaCl pH 8, for 30 min at room temperature in the dark. FITC-labeled bacteria were wash 2-3 times and suspended in PBS at $1\text{-}1.5x10^9$/ml. 200µl of this suspension ($2\text{-}3x10^8$) were incubated with 200µl ($1x10^5$) epithelial cells for 30min a 37°C. Cells were than centrifuged at 2000rpm for 5 min to remove non-adherent bacteria, suspended in 200µl of PBS, transferred to FACScan tubes and read

CLAIMS

1. A method for the heterologous expression of a protein of the invention, in which (a) at least one domain in the protein is deleted and, optionally, (b) no fusion partner is used.

2. The method of claim 1, in which the protein of the invention is ORF46.

3. The method of claim 2, in which ORF46 is divided into a first domain (amino acids 1-433) and a second domain (amino acids 433-608).

4. The method of claim 2, in which the protein of the invention is 564.

5. The method of claim 4, in which protein 564 is divided into domains as shown in Figure 8.

6. The method of claim 1 in which the protein of the invention is 961.

7. The method of claim 6, in which protein 961 is divided into domains as shown in Figure 12.

8. The method of claim 1, in which the protein of the invention is 502 and the domain is amino acids 28 to 167 (numbered according to the MC58 sequence).

9. The method of claim 1, in which the protein of the invention is 287.

10. A method for the heterologous expression of a protein of the invention, in which (a) a portion of the N-terminal domain of the protein is deleted.

11. The method of claim 9 or claim 10, in which protein 287 is divided into domains A B & C shown in Figure 5.

12. The method of claim 11, in which (i) domain A, (ii) domains A and B, or (iii) domains A and C are deleted.

13. The method of claim 11, wherein (i) amino acids 1-17, (ii) amino acids 1-25, (iii) amino acids 1-69, or (iv) amino acids 1-106, of domain A are deleted.

14. A method for the heterologous expression of a protein of the invention, in which (a) no fusion partner is used, and (b) the protein's native leader peptide (if present) is used.

15. The method of claim 14, in which the protein of the invention is selected from the group consisting of: 111, 149, 206, 225-1, 235, 247-1, 274, 283, 286, 292, 401, 406, 502-1, 503, 519-1, 525-1, 552, 556, 557, 570, 576-1, 580, 583, 664, 759, 907, 913, 920-1, 936-1, 953, 961, 983, 989, Orf4, Orf7-1, Orf9-1, Orf23, Orf25, Orf37, Orf38, Orf40, Orf40.1, Orf40.2, Orf72-1, Orf76-1, Orf85-2, Orf91, Orf97-1, Orf119, Orf143.1, NMB0109, NMB2050, 008, 105, 117-1, 121-1, 122-1, 128-1, 148, 216, 243, 308, 593, 652, 726, 926, 982, Orf83-1 and Orf143-1.

16. A method for the heterologous expression of a protein of the invention, in which (a) the protein's leader peptide is replaced by the leader peptide from a different protein and, optionally, (b) no fusion partner is used.

17. The method of claim 16, in which the different protein is 961, ORF4, *E.coli* OmpA, or *E.carotovora* PelB, or in which the leader peptide is MKKYLFSAA.

18. The method of claim 17, in which the different protein is *E.coli* OmpA and the protein of the invention is ORF1.

19. The method of claim 17, in which the protein of the invention is 911 and the different protein is *E.carotovora* PelB or *E.coli* OmpA.

20. The method of claim 17, in which the different protein is ORF4 and the protein of the invention is 287.

21. A method for the heterologous expression of a protein of the invention, in which (a) the protein's leader peptide is deleted and, optionally, (b) no fusion partner is used.

22. The method of claim 21, in which the protein of the invention is 919.

23. A method for the heterologous expression of a protein of the invention, in which expression of a protein of the invention is carried out at a temperature at which a toxic activity of the protein is not manifested.

24. The method of claim 23, in which protein 919 is expressed at 30°C.

25. A method for the heterologous expression of a protein of the invention, in which protein is mutated to reduce or eliminate toxic activity.

26. The method of claim 25, in which the protein of the invention is 907, 919 or 922.

27. The method of claim 26, in which 907 is mutated at Glu-117 (*e.g.* Glu→Gly).

28. The method of claim 26, in which 919 is mutated at Glu-255 (*e.g.* Glu→Gly) and/or Glu-323 (*e.g.* Glu→Gly).

29. The method of claim 26, in which 922 is mutated at Glu-164 (*e.g.* Glu→Gly), Ser-213 (*e.g.* Ser→Gly) and/or Asn-348 (*e.g.* Asn→Gly).

30. A method for the heterologous expression of a protein of the invention, in which vector pSM214 is used or vector pET-24b is used.

31. The method of claim 30, in which the protein of the invention is 953 and the vector is pSM214.

32. A method for the heterologous expression of a protein of the invention, in which a protein is expressed or purified such that it adopts a particular multimeric form.

33. The method of claim 32, in which protein 953 is expressed and/or purified in monomeric form.

34. The method of claim 32, in which protein 961 is expressed and/or purified in tetrameric form.

35. The method of claim 32, in which protein 287 is expressed and/or purified in dimeric form.

36. The method of claim 32, in which protein 919 is expressed and/or purified in monomeric form.

37. A method for the heterologous expression of a protein of the invention, in which the protein is expressed as a lipidated protein.

38. The method of claim 37, in which the protein of the invention is 919, 287, ORF4, 406, 576, or ORF25.

39. A method for the heterologous expression of a protein of the invention, in which (a) the protein's C-terminus region is mutated and, optionally, (b) no fusion partner is used.

40. The method of claim 39, wherein the mutation is a substitution, an insertion, or a deletion

41. The method of claim 40, wherein the protein of the invention is 730, ORF29 or ORF46.

42. A method for the heterologous expression of a protein of the invention, in which the protein's leader peptide is mutated.

43. The method of claim 42, in which the protein of the invention is 919.

44. A method for the heterologous expression of a protein, in which a poly-glycine stretch within the protein is mutated.

45. The method of claim 44, wherein the protein is a protein of the invention.

46. The method of claim 45, wherein the protein of the invention is 287, 741, 983 or Tbp2.

47. The method of claim 46, wherein $(Gly)_6$ is deleted from 287 or 983.

48. The method of claim 46, wherein $(Gly)_4$ is deleted from Tbp2 or 741

49. The method of claim 47 or claim 48, wherein the leader peptide is also deleted.

50. The method of any preceding claim, in which the heterologous expression is in an *E.coli* host.

51. A protein expressed by the method of any preceding claim.

52. A heterologous protein comprising the N-terminal amino acid sequence MKKYLFSAA.

FIGURE 1

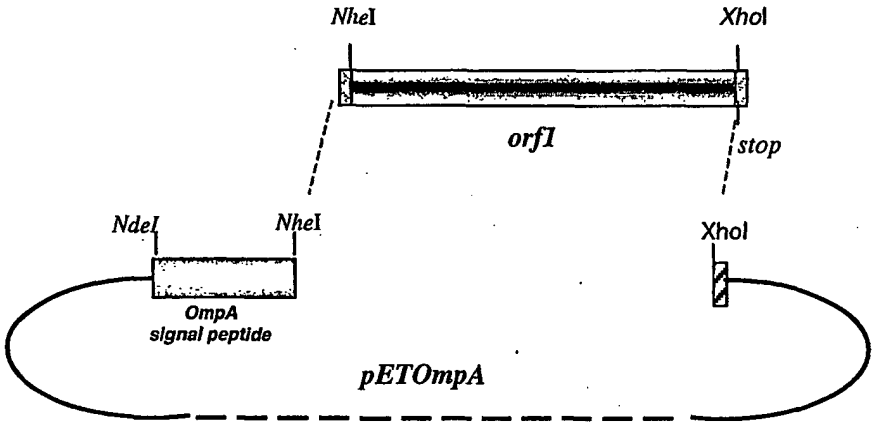


FIGURE 2

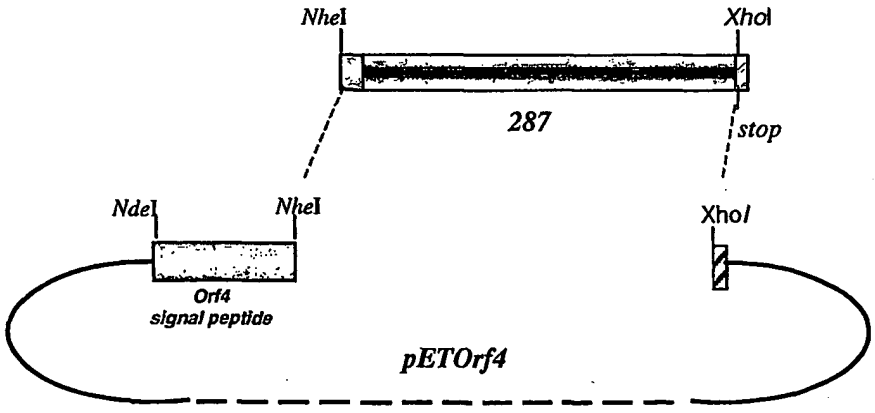
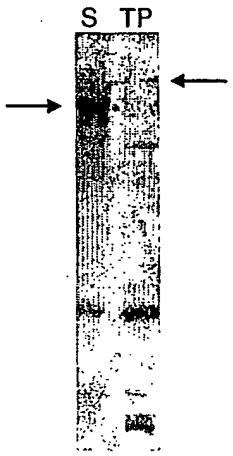
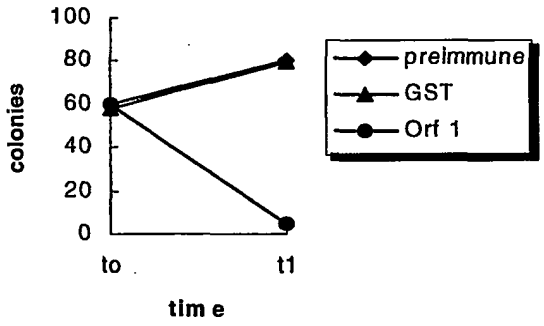


FIGURE 3

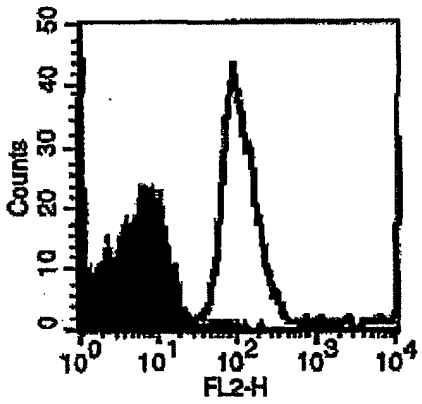
PURIFICATION



WESTERN BLOT



BACTERICIDAL ASSAY

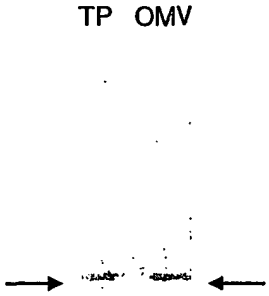


FACS

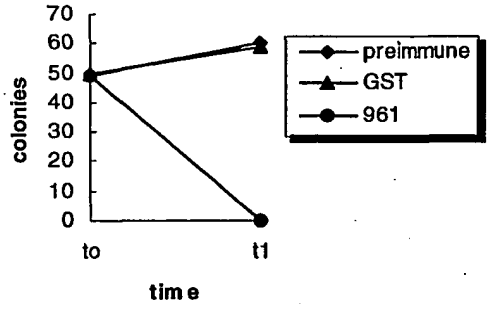
ELISA: *POSITIVE*

FIGURE 4

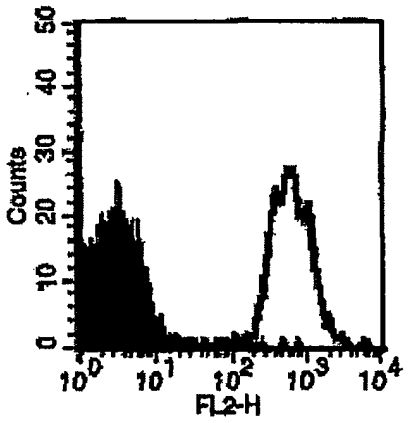
PURIFICATION



WESTERN BLOT



BACTERICIDAL ASSAY



FACS

ELISA: *POSITIVE*

FIGURE 5

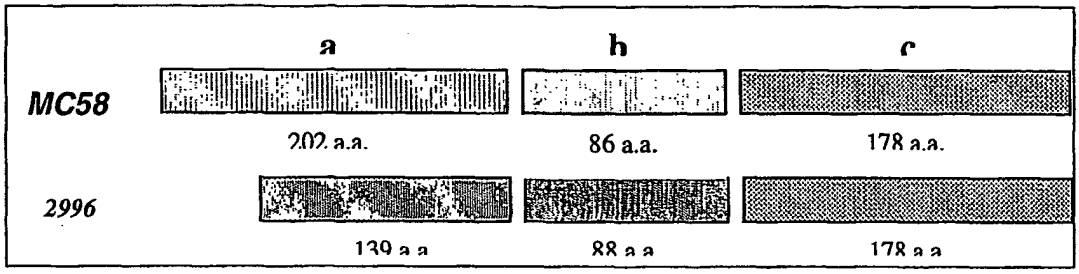


FIGURE 6

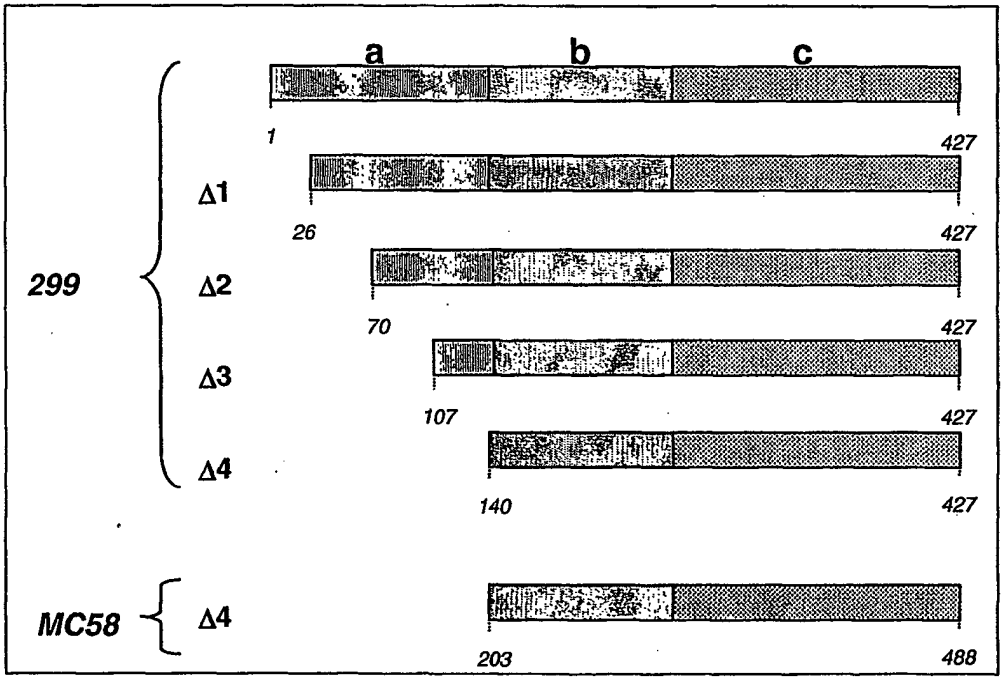


FIGURE 9

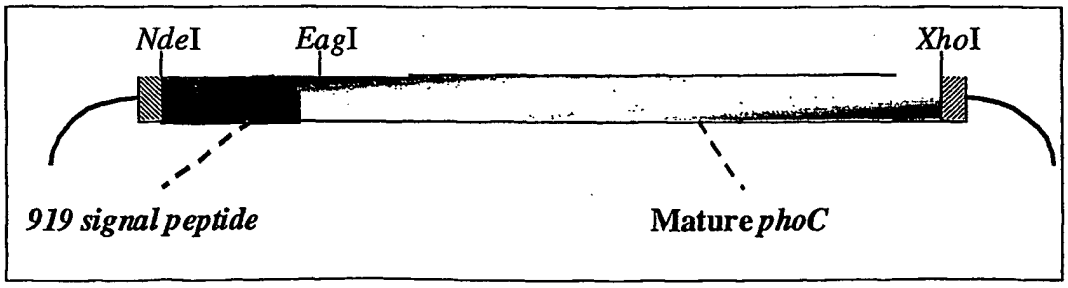


FIGURE 7

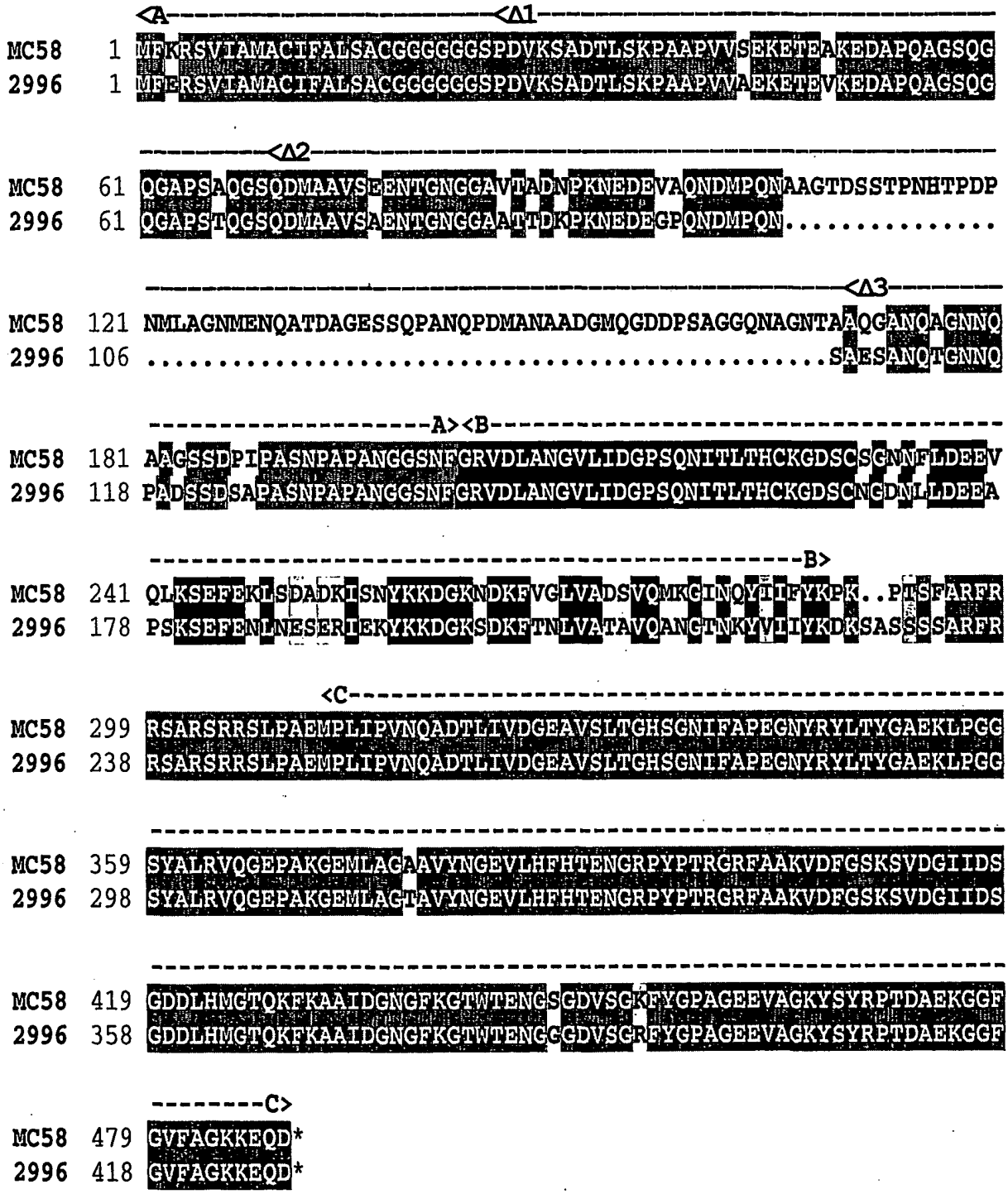

<A
<Δ1
MC58 1 MFKRSVIAMACIFALSACGGGGGGSPDVKSADTLSKPAAPVVSEKETEAKEDAPQAGSQG
2996 1 MFERSVIAMACIFALSACGGGGGGSPDVKSADTLSKPAAPVVAEKETEVKEDAPQAGSQG
<Δ2
MC58 61 QGAPSAQGSQDMAAVSEENTGNGGAVTADNPKNEDEVAQNDMPQNAAGTDSSTPNHTPDP
2996 61 QGAPSTQGSQDMAAVSAENTGNGGAATTDKPKNEDEGPQNDMPQN.....
<Δ3
MC58 121 NMLAGNMENQATDAGESSQPANQPDMANAADGMQGDDPSAGGQNAGNTAAQGANQAGNNQ
2996 106 .....SAESANQTGNNQ
A><B
MC58 181 AAGSSDPIPASNPAPANGGSNFGRVDLANGVLIDGPSQNITLTHCKGDSCSGNNFLDEEV
2996 118 PADSSDSAPASNPAPANGGSNFGRVDLANGVLIDGPSQNITLTHCKGDSCNGDNLLDEEA
B>
MC58 241 QLKSEFEKLSDADKISNYKKDGKNDKFVGLVADSVQMKGINQYIIFYKPK..PTSFARFR
2996 178 PSKSEFENLNESERIEKYKKDGKSDKFTNLVATAVQANGTNKYVLIYKDKSASSSSARFR
<C
MC58 299 RSARSRRSLPAEMPLIPVNQADTLIVDGEAVSLTGHSGNIFAPEGNYRYLTYGAEKLPGG
2996 238 RSARSRRSLPAEMPLIPVNQADTLIVDGEAVSLTGHSGNIFAPEGNYRYLTYGAEKLPGG
MC58 359 SYALRVQGEPAKGEMLAGAAVYNGEVLHFHTENGRPYPTRGRFAAKVDFGSKSVDGIIDS
2996 298 SYALRVQGEPAKGEMLAGTAVYNGEVLHFHTENGRPYPTRGRFAAKVDFGSKSVDGIIDS
MC58 419 GDDLHMGTQKFKAAIDGNGFKGTWTENGSGDVSGKFYGPAGEEVAGKYSYRPTDAEKGGF
2996 358 GDDLHMGTQKFKAAIDGNGFKGTWTENGCGDVSGRFYGPAGEEVAGKYSYRPTDAEKGGF
C>
MC58 479 GVFAGKKEQD*
2996 418 GVFAGKKEQD*

FIGURE 8

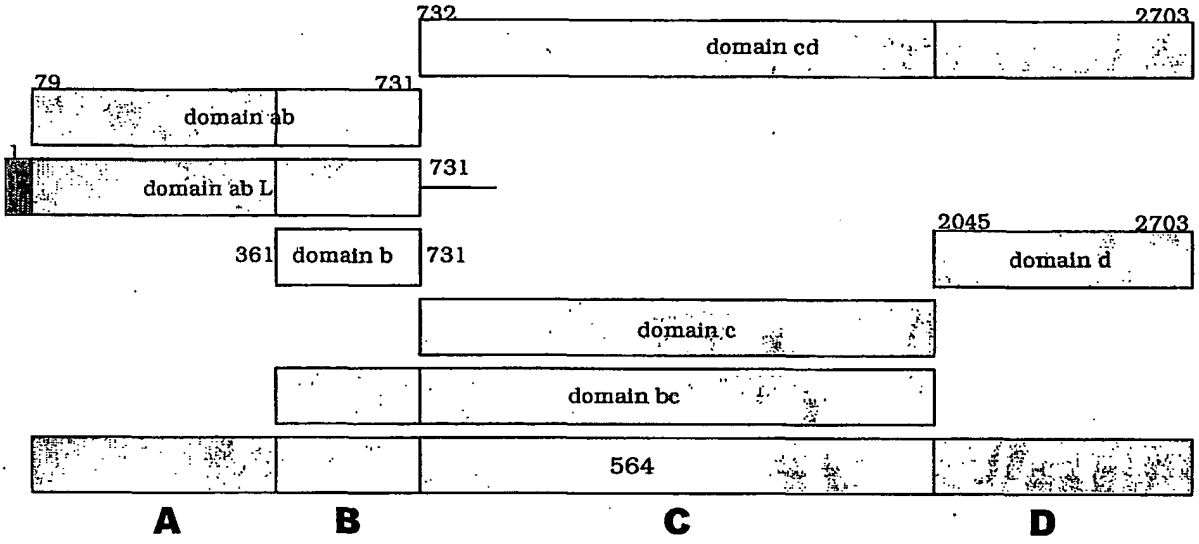


FIGURE 10

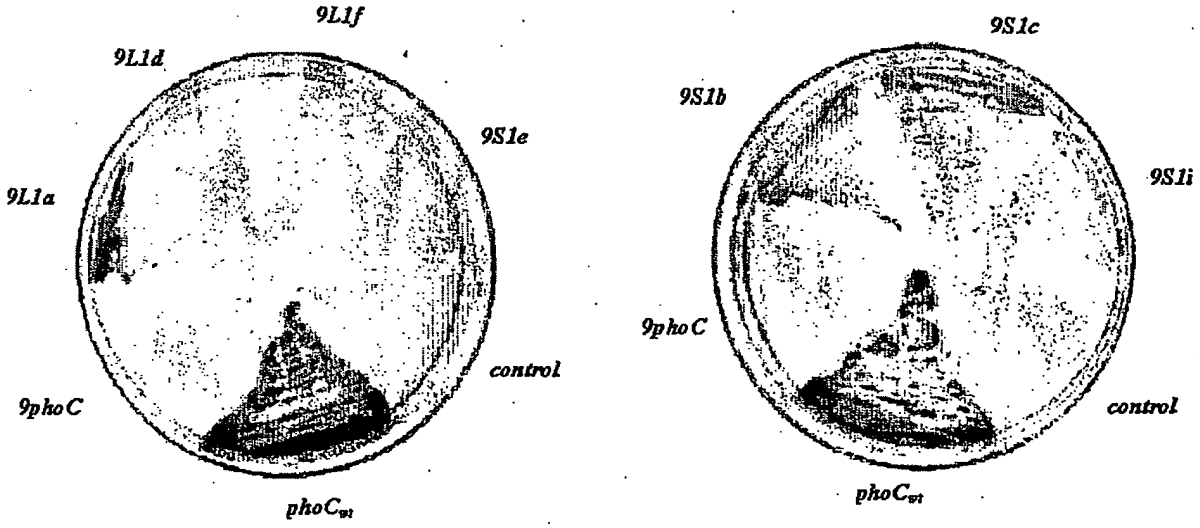


FIGURE 11A

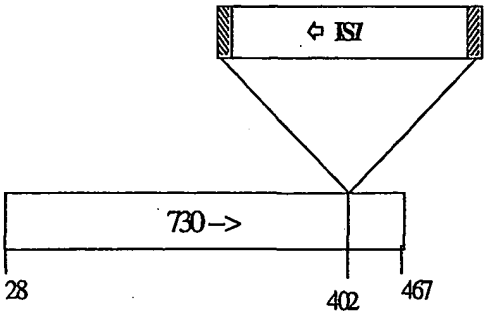


FIGURE 11B

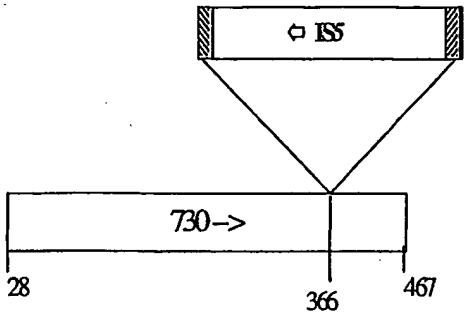
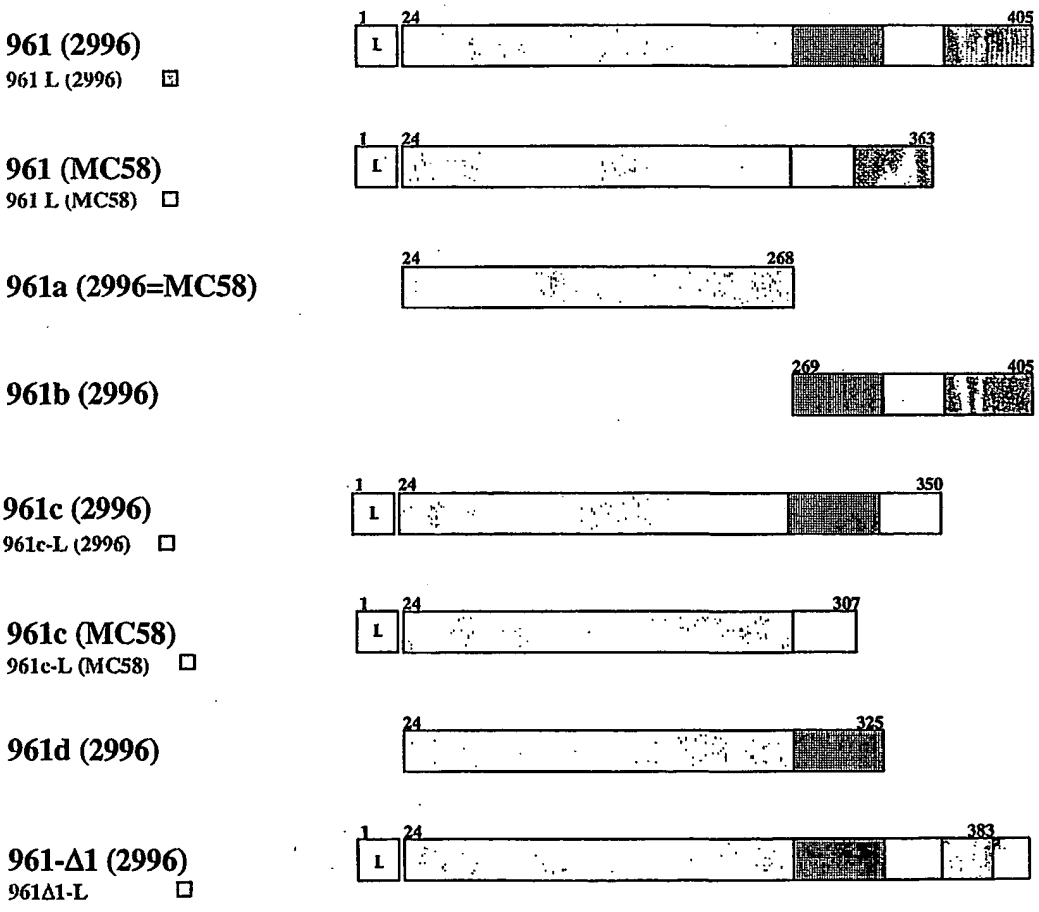


FIGURE 12



Leader Peptide | Region present in 2996, not in MC58 | Coil-coiled segment | Membrane anchor

FIGURE 13

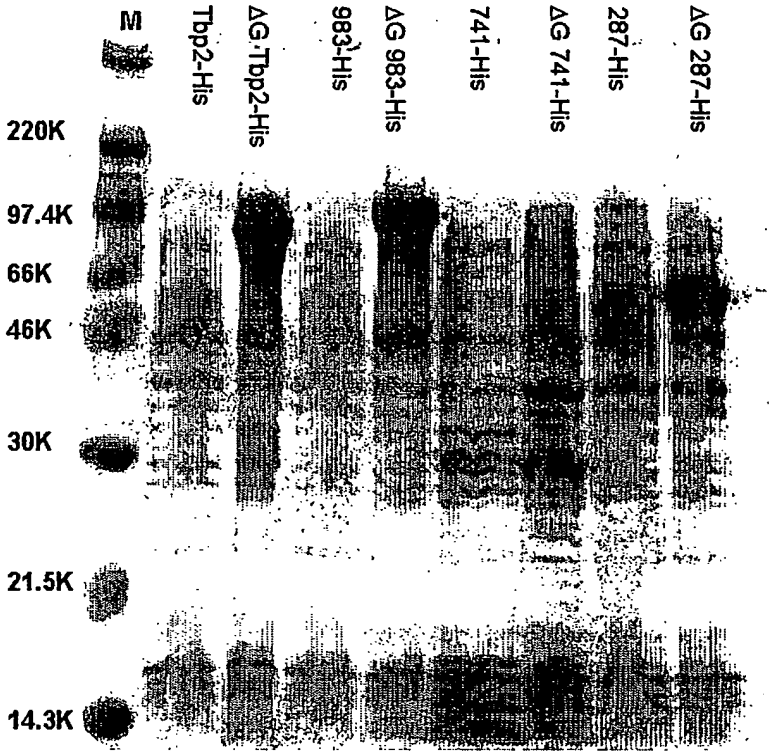


FIGURE 14

FIGURE 14A — ΔG287—919

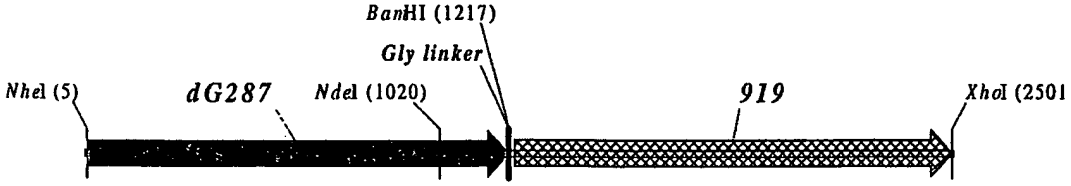


FIGURE 14B — ΔG287—953

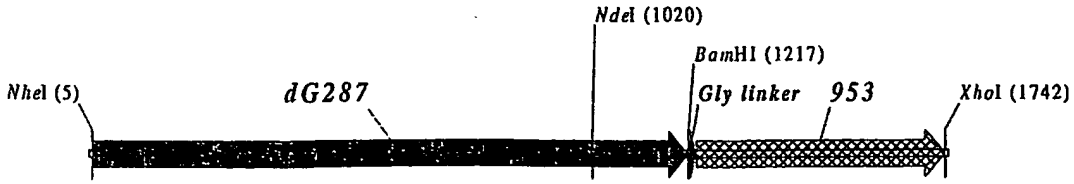


FIGURE 14C — ΔG287—961

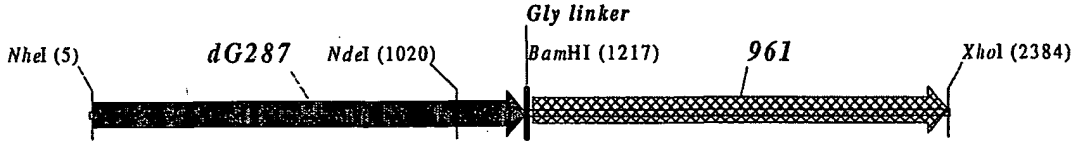


FIGURE 14D — ΔG287NZ—919

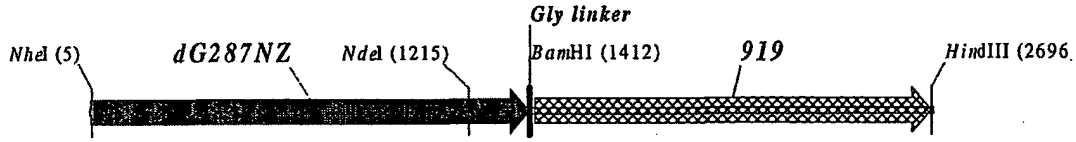


FIGURE 14E — ΔG287NZ—953

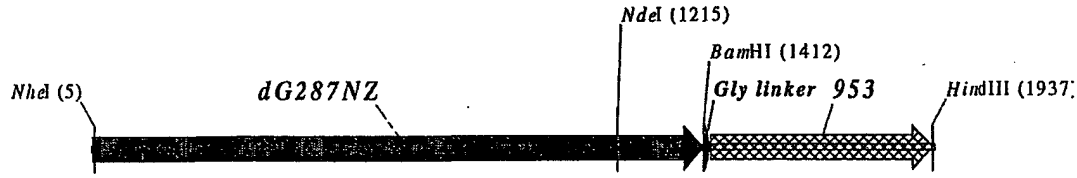


FIGURE 14F — ΔG287NZ—961

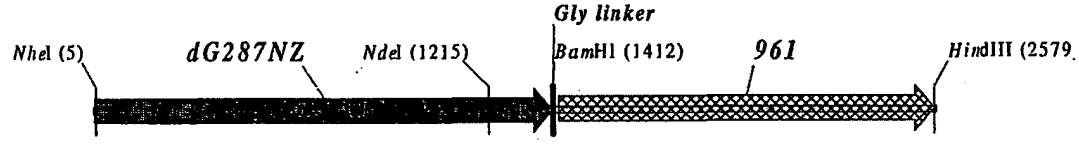


FIGURE 14G — ΔG983-ORF46.1

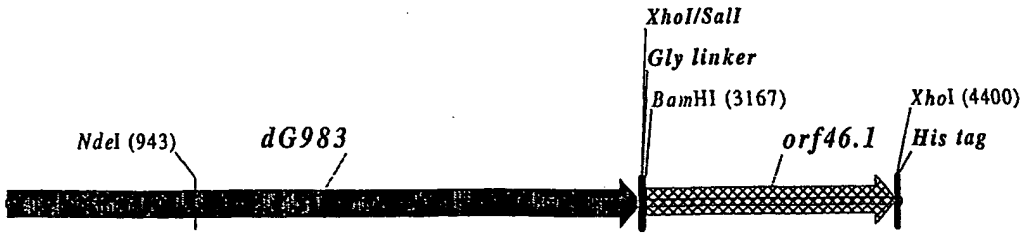


FIGURE 14H — ΔG983-741

FIGURE 14I — ΔG983-961

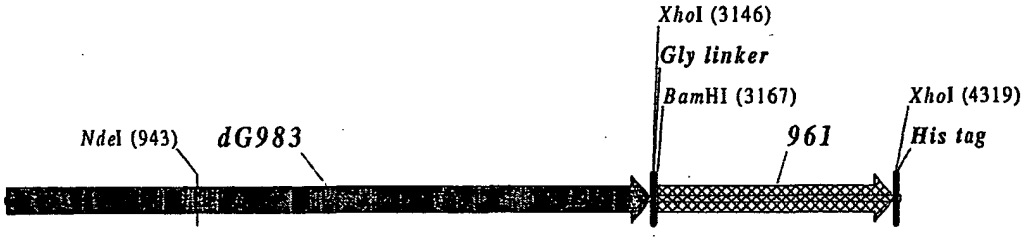


FIGURE 14J — ΔG983-961c

FIGURE 14K — ΔG741-961

FIGURE 14L — ΔG741-961c

FIGURE 14M — ΔG741-983

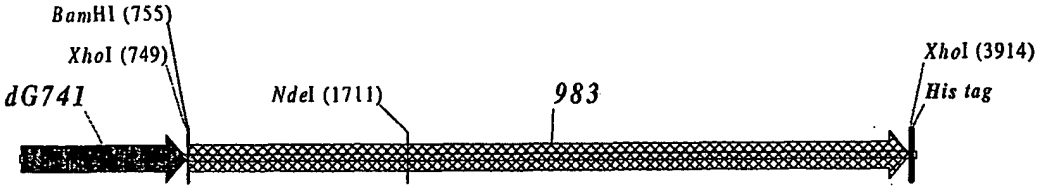


FIGURE 14N — ΔG741-ORF46.1

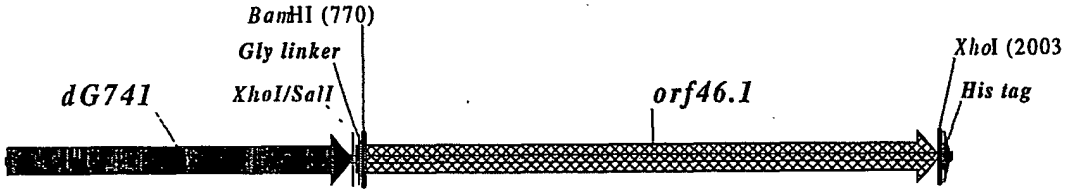


FIGURE 14O — ORF46.1-741

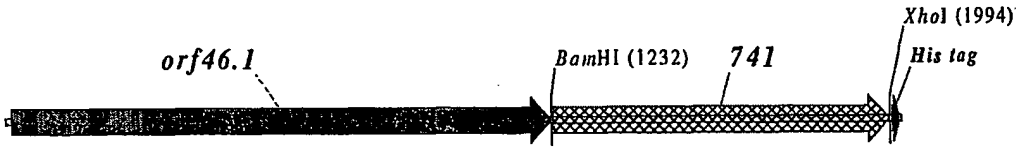


FIGURE 14P — ORF46.1-961

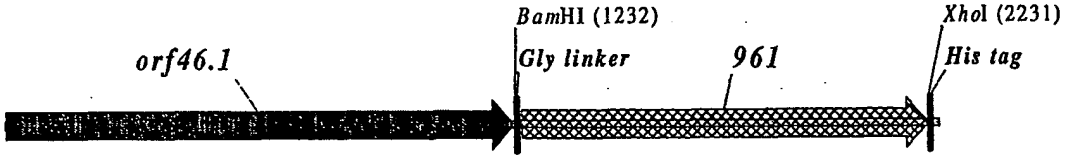


FIGURE 14Q — ORF46.1—961c

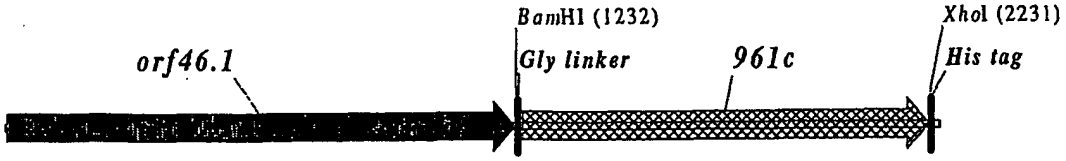


FIGURE 14R — 961-ORF46.1

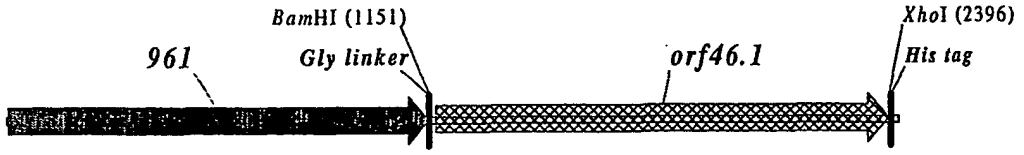


FIGURE 14S — 961-741

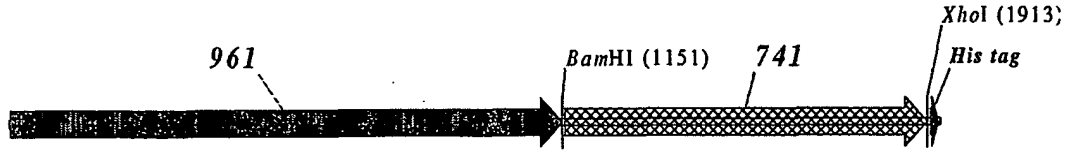


FIGURE 14T — 961-983

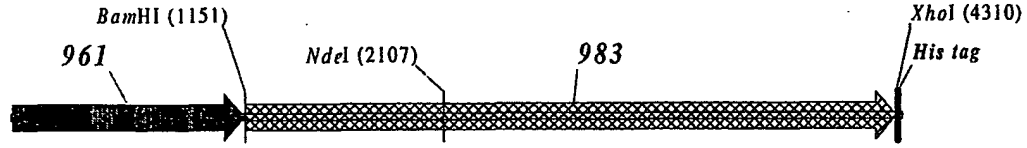


FIGURE 14U — 961c-ORF46.1

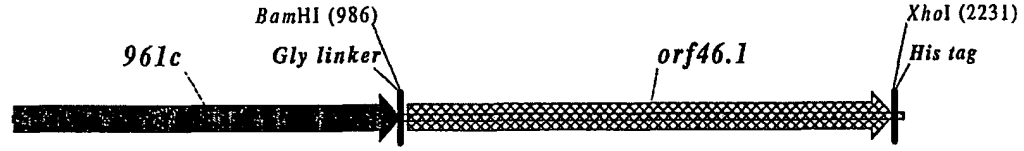


FIGURE 14V — 961c-741

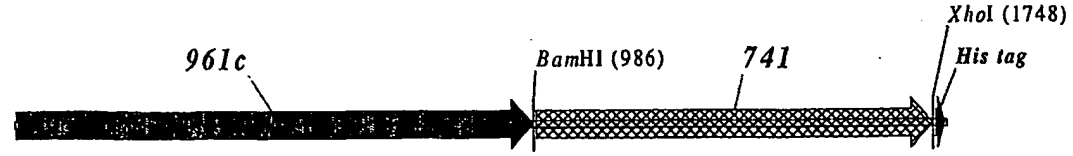


FIGURE 14W — 961c-983

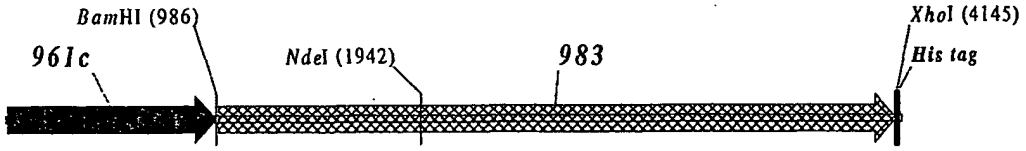


FIGURE 14X — 961cL-ORF46.1

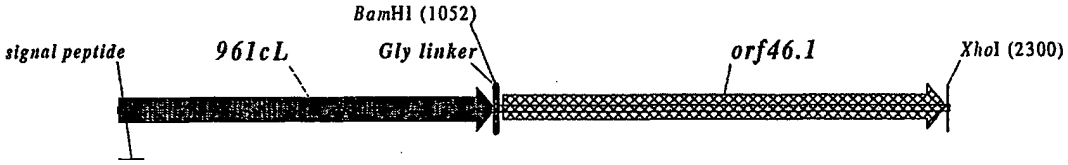


FIGURE 14Y — 961cL-741

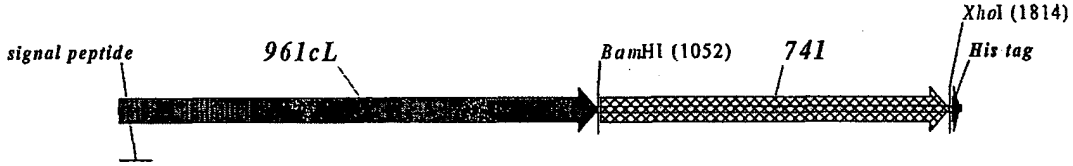


FIGURE 14Z — 961cL-983